

The centre needs to hold

Research ministers met for a seminar in London this week to discuss how to improve the handling of European-level research. The solution lies in steady and targeted reforms to practices in Brussels.

The relative lack of opposition with which many Europeans appear to have accepted the imminent introduction of a single currency — the unimaginatively named Euro — has surprised even some of its strongest critics. The general mood seems to reflect a passive acceptance that, like democracy itself, it may not be a perfect way of running things, but it is the best option we know of.

A similar judgement might be made about the management of the joint research programmes of the European Union. There are many valid criticisms of the way Brussels handles research funds. Excessive form filling, the need to meet obscure bureaucratic requirements, lengthy delays in reaching decisions, high rejection rates, excessive political interference, and remoteness from national research priorities and the activities of national research agencies; almost every European researcher who has applied for such funds, whether successful or not, has his or her own horror story to tell.

Unsurprisingly, such complaints have encouraged a search for alternative, sometimes radical, solutions. Some have been arguing, with considerable logic on their side, that more responsibility for the management of research programmes should be contracted out by the European Commission to the research community through bodies such as the European Science Foundation. Others believe that national organizations should be allowed to play a greater role in helping to manage European-level funds. Perhaps most provocatively, the French research minister, Claude Allègre, has recently been suggesting a full-scale devolution of some management responsibilities from Brussels to national agencies, such as France's own Centre Nationale de la Recherche Scientifique.

Some movement in these directions would undoubtedly be beneficial. Centralized bureaucracies can easily become isolated from those they are intended to serve. And experiments such as the AMICA consortium, which has been given responsibility for allocating research funds in the field of plant biology, have been of substantial value. But an excessive devolution of responsibility away from Brussels would be a mistake at a time when Europe needs more, not

less, strategic thinking about the development of its research base.

Take, for example, the European Union's programme on Training and Mobility of Researchers. Suggestions have been heard recently that responsibility for fellowships awarded through this programme should be passed back to national organizations able to handle them in a more 'rational' manner. That might work in countries with well-organized national institutions for handling such schemes. For others who prefer to work in a less centralized way — and who find the current arrangement largely satisfactory, particularly in the way it allows a genuinely European level of evaluation and assessment — such dramatic change holds little attraction.

Not even the most hardened commissioner would argue that Brussels is perfect. The agenda of a top-level seminar attended by almost a dozen research ministers in London this week (see page 849) was replete with topics on which the commission could do better. These range from the trivial but irritating — such as why, in an age of electronic communication, so many application forms still require to be painstakingly typed — to strategic issues about ensuring that the research projects in the forthcoming Fifth Framework Programme genuinely reflect the interests of potential users of such research.

There are many ways in which the commission needs shaking up in its handling of research; by concentrating the minds of his colleagues on the Council of Ministers on some of these, Allègre has already performed a useful service. But many in the commission are well aware of such needs, and are striving to address them. Critics also sometimes ignore the complex political pressures under which the commission is required to operate. Even researchers, for example, appreciate that applying the concept of 'cohesion' may not produce the most cost-effective science. But it has other advantages in terms of disseminating scientific and technical expertise. There are times when deliberate, well-planned reform can achieve much more than revolutionary rhetoric. This is what Brussels now needs; not ambitious, and sometimes misguided, claims that others could do its job better. □

From art to image

As a series of articles changes direction, science's relevance to art is reaffirmed.

This issue sees the concluding article in the first part of a series by the art historian Martin Kemp. In previous articles, he has dealt with scientific ideas and perspectives embedded in 27 artists' handling of their subject matter. Such an approach is illuminating provided one understands the science (as Kemp clearly does), the art is worth looking at and analysing (as at least most of it has been), and especially where a writer delivers a highly focused presentation (as is inevitably required in a one-page format). On pages 875–876, and at slightly greater length, Kemp places much of what he has discussed into a broader summarizing framework. Many readers will no doubt have already formed their own judgements. But the result of the series has surely been at least a broadened range of readers' awareness of artists and a deepened appreciation of how art works, as well as lively discussions in other fora, such as museums and galleries.

The series now shifts to the same art historian's perspective on the way scientific concepts and objects are visualized, starting next week with a discussion of the use of images in this journal. Kemp will subsequently deal with scientific images with the same three-week cyclical approach as previously: pre-twentieth century, classic twentieth century, contemporary.

Physicists in America, nervous about pseudoscience, are worrying about a proposed public statement (see page 849) which in draft form emphasizes science's objective and repeatable qualities, while (more controversially) welcoming complementary approaches to the understanding of nature. Kemp's series has illuminated artistic subjectivity and objectivity. In doing so, however, he has shown not that art assists scientific understanding, but that scientific insight can provide a fruitful stimulus to some of the most powerful of artistic interpretations. □



'More effectiveness needed' in Brussels' handling of research

KEITH DOBNEY

[LONDON] Research ministers from member states of the European Union (EU) have been asked to consider a set of measures proposed this week by the European Commission in Brussels to improve the effectiveness with which research programmes are managed.

One of the commission's suggestions — drafted in response to pressure from EU member states for a more imaginative and innovative approach to research management — is a regular 'benchmarking' exercise comparing its performance against that of other research organizations.

The ministers also want those actively engaged in research to be more involved in setting up research programmes. And the commission is proposing a pilot experiment on how the management of grants aimed to increase the mobility of research workers in Europe might be devolved from Brussels, giving more autonomy to host institutions.

The proposals were submitted on Tuesday 28 April by Edith Cresson, the commissioner responsible for research and education, to a colloquium in London attended by more than half of the 15 EU research ministers.



Battle: keen to see greater user input



Cresson: accepting pressure for change

Although hosted by the British government — which holds the chair of the Council of Ministers — the meeting was held primarily at the suggestion of Claude Allègre, the French minister for research, technology and education.

It reflected a feeling that, now that the budget of the Fifth five-year Framework research programme (FP5) has been approved (see *Nature* 391, 729; 1998), the time is ripe to turn a more critical eye on how effectively the programme is managed.

Allègre, for example, has suggested ways in which responsibility for day-to-day management might be devolved away from Brussels to organizations in member states. Other changes in the organization of European-level research have been discussed among the heads of the research councils of EU member states, who meet regularly as the group known as Eurohorcs.

The day before the London meeting, Eurohorc representatives met European Science Foundation (ESF) officials in Strasbourg to discuss how the two bodies might work together to advise the commission on the use of EU research funds.

No decisions were taken at the meeting. It was agreed, however, that the ESF (whose members also include non-EU states) would explore how such closer collaboration might work in practice; for example, whether it should set up a small secretariat for Eurohorcs in its Strasbourg offices.

Brussels officials are said to be watching such developments closely, aware both of the scope for more effective management of individual programmes through devolution, but also of the danger of diluting the 'added value' which, they argue, comes from operating through the commission. Some are also said to be concerned that devolution might reduce their own authority.

Prior to Tuesday's meeting, John Battle, Britain's science minister, said that the ECU14 billion (\$15.5 billion) for FP5 represented "a huge amount of money which must be managed carefully and transparently".

Some of the ideas being discussed reflect the general approach to research management that the British government is keen to see adopted in Brussels.

The idea of 'benchmarking' performance by comparing it with that in other countries, for example, has attracted a strong following both in the Office of Science and Technology and in university funding councils.

Indeed, the commission itself has already responded to such interest by emphasizing, in a statement issued on Monday, that a recent study by Andersen Consulting shows that its administrative costs "compare favourably [original emphasis] with those of other European and national research organizations".

Following Tuesday's meeting, the commission will put forward proposals on how it intends to respond to the ministers' demands for more effective management of research funds at the next formal meeting of the EU research council in Brussels on 22 June.

David Dickson & Alison Abbott

Physicists seek definition of 'science'

[WASHINGTON] The governing council of the American Physical Society (APS) has rejected the first draft of a statement defining science for the public, which the society's public affairs panel has been preparing for three years. According to an official familiar with the discussion, some members were concerned by a proposed reference to "other approaches" to understanding nature.

Others are said to have been worried about public misunderstanding of the statement's references to "falsifiability". The authors of the draft 200-word statement have been asked to confer with scientific societies and other interested parties before coming back with a new version later this year.

The case for such a statement has been recently confirmed by opinion polls showing that public belief in forms of pseudoscience — such as faith healing and astrology — is growing in the United States. But the rejection of the draft, although not unusual for such a policy statement, illustrates the difficulties that scientists face in trying to draw a recognizable line between their own work and pseudoscience.

The statement, entitled "What is Science?", defines science as "a disciplined quest to understand nature in all its aspects"

and explains that it demands both "open and complete exchange of ideas and data" and "an attitude of scepticism about its own tenets".

It stresses that scientific results must be capable of reproduction, modification or falsification by independent observers. And it closes by noting that "scientists value other, complementary approaches to and methods of understanding nature", but that "if the alternatives are to be called scientific, they must adhere to the principles outlined above".

Following the draft's rejection by the council at a meeting last week in Columbia, Ohio, the APS may now draw up two statements — one for wide public dissemination and the other a more rigorous explanatory statement for scientists themselves.

The society decided to produce the statement in response to the concerns of key members that 'pseudoscience' is not only winning increased public attention but may even be causing confusion among science students. APS members have been active in criticizing this trend not just in cases related to physics — such as the alleged discovery of 'cold fusion' — but also in other fields, such as alternative medicine.

Colin Macilwain

UK waives nuclear waste rule for Georgia

[MUNICH] Britain has temporarily waived an agreement that reprocessed nuclear fuel must be returned to its country of origin to enable it to accept a consignment of fresh and spent fuel from Georgia.

The 5.1-kg consignment, from a shut-down research reactor near Tbilisi, was transferred to a reprocessing facility at Dounreay in northern Scotland last week. The move, which has been harshly criticized by some anti-nuclear activists, was designed to reduce the risk of illegal trade in weapons-grade uranium.

Critics are angry that the agreement to import the fuel, reached between British Prime Minister Tony Blair and US President Bill Clinton, was not debated publicly, and did not take into account recent concerns about safety at the Dounreay reprocessing plants, which are at present standing idle.

Only hours before the government's announcement, the Nuclear Installations Inspectorate (NII), part of the UK government's Health and Safety Executive, had announced a ban on the import of radioactive material for reprocessing until a complete safety audit and an assessment of

the plant's waste-management capability had been carried out.

In recent years, fragments of highly enriched uranium have been found on beaches near the facility, and at least one serious leak — in the dissolver tank of its fast reprocessing plant — occurred in 1996. Bringing the Dounreay plants to a safety level acceptable to the NII could take two years.

Most of the Georgian consignment is fresh fuel, which the UK Foreign Office says will be used up as raw material in Dounreay's facilities for producing medical isotopes. Only 0.8 kg is spent fuel requiring reprocessing; this will be stored on site until Dounreay gets the green light to recommence reprocessing.

Since the collapse of the Soviet Union, the nuclear powers have been sharing the responsibility for nuclear safety. The United States, for example, has taken 600 kg of highly enriched uranium (HEU) out of Kazakhstan, while Canada, Germany and France have helped tighten up safety at nuclear reactors.

"There is certainly a feeling that Britain should share some of the burden," says a UK government spokesman, arguing that there has been a "considerable overreaction to

the decision to take in such a small amount of fuel".

The research reactor from which the fuel originates is at the Georgia Institute of Physics — near Tbilisi, Georgia's capital — which stopped operations after the Chernobyl disaster in 1986.

During the Soviet era, Georgia could ship its spent fuel to the Russian nuclear complex at Chelyabinsk in the Ural mountains. But in 1991 Russia ended the arrangement because, following the collapse of the Soviet Union, Georgia's status as a foreign country meant it was no longer eligible to use the Urals site.

A spokesman for the environmental group Greenpeace International, Mike Townsley, reluctantly accepts the "superficial logic" of bringing the Georgian fuel to Britain, despite concerns about safety at Dounreay, in view of bigger nuclear proliferation issues. "Given the choice between terrorists and 'mad scientists', it is obviously better to come down on the side of 'mad scientists'," he says.

"But more importantly, the Georgia affair is a wake-up call for Britain to the general dangers of the movement of weapons-grade nuclear material which is scattered around the world." Greenpeace International is calling for an immediate full and open debate on the issue.

Britain's own rules for reprocessing foreign radioactive waste require that reprocessed fuel be returned to the country of origin for long-term storage, to prevent Britain becoming a 'nuclear dumping ground'. But this rule was waived on a 'one-off' basis for the Georgian consignment in the interests of non-proliferation, says a spokesman for the Health and Safety Executive.

US rules for supplying HEU to research reactors require that the nuclear waste be returned to the United States for storage and reprocessing, to prevent commerce in bomb-grade uranium. As a consequence, Dounreay has had difficulties finding customers for its reprocessing facilities, and had to put its main reprocessing plant on idle in 1996.

But Dounreay had hoped to start reprocessing again soon, after recently winning two large contracts to reprocess spent HEU fuel rods from research reactors. One is with ICI in England, which had originally received its HEU fuel from Russia; handling its waste will require Dounreay to obtain permission to open its new reprocessing plant designed for aluminium-clad fuel.

The other contract is with the Australian Nuclear Science and Technology Organization near Sydney (see *Nature* 389, 109; 1997), whose research reactor was supplied with HEU fuel by Britain in the 1960s. Both the Australian body and ICI have been informed that Dounreay cannot at present accept their fuel.

Alison Abbott

German reactor project faces safety challenge

[MUNICH] A US nuclear watchdog organization has called for the construction of a research reactor at a German university to be suspended on the grounds that the reactor's proposed fuel has been inadequately tested for safety.

The Technical University of Munich had planned to test a sample rod of the fuel — densely packed highly enriched uranium (HEU) — from the FRMII reactor in the SILOE reactor in Grenoble, France. But the university abandoned the idea when it became clear that the amount of heat generated would exceed the limits of the reactor.

Instead, university researchers tested a fuel rod of intermediate density — 1.5 grams of uranium per cubic centimetre instead of 3.0 grams — and extrapolated the results to the high-density fuels. This extrapolation was accepted by the Technische Überwachungs Verein, the German atomic-plant

inspectorate that advises licensing authorities, to approve the building of the reactor, which began in 1996.

The Washington-based Nuclear Control Institute (NCI) has long argued that the reactor should not have been designed to burn weapons-grade HEU. Under a supplementary agreement to the Nuclear Non-Proliferation Treaty, signatories agreed to convert existing HEU-burning research reactors to low-enriched uranium (LEU), and not to build new HEU-burning reactors.

The NCI has written to the Bavarian environment minister, who will be responsible for licensing the operation of the reactor, saying that without complete tests "it is impossible for a licensing authority to certify the safety of a proposed reactor".

But the university insists that the safety tests were sufficient and in accordance with international standards. "We have conducted tests on

those parts of the fuel plates which have the highest uranium fission density," says Gert von Hassel, spokesman for the FRMII project. Thus, he says, extrapolation to HEU fuel rods is legitimate.

The NCI says that the risks of catastrophic accident and nuclear terrorism could be averted if the reactor were redesigned to burn low-enriched uranium fuel. A recent study by the Argonne National Laboratory (ANL) in the United States indicated that the reactor could produce the same experimental performance if it converted. But according to von Hassel, the ANL proposal is limited to fuel consideration and says nothing about safety.

The reactor could still be forced back to the drawing-board if the opposition Social Democrat Party wins power in the September federal elections. The party wants an alternative reactor, based on the LEU, to be developed.

A. A. & Quirin Schiermeier

Jewish leaders meet NIH chiefs on genetic stigmatization fears

[WASHINGTON] Senior officials from the National Institutes of Health (NIH) last week met publicly for the first time with a broad spectrum of Jewish leaders on the increasingly sensitive issue of the meaning of genetics research for Jews (see *Nature* 389, 322; 1997).

Francis Collins, director of the National Human Genome Research Institute and Richard Klausner, director of the National Cancer Institute, addressed 40 Jewish leaders covering denominations from ultra-liberal to Orthodox in a meeting in Washington.

Marlene Post, president of Hadassah, the Jewish women's organization that sponsored the meeting with the Jewish Council for Public Affairs, said that all participants shared a concern that fear of discrimination and stigmatization "could actually stem important genetic research". It might also discourage individuals from seeking important health information through genetic testing.

"We need to protect ourselves both from discrimination and ignorance," Klausner told the meeting. "We've got to start breaking down myths that ethnic groups mean genetic groups. They don't."

Both the scientists and Jewish participants expressed concern over the lack of legal protections against genetic discrimination. The US Congress has not passed any comprehensive law against job or insurance discrimination based on genetic information, and appears highly unlikely to do so this year.

There are also signs that fears of stigmatization and of losing, or being denied, jobs and health insurance are preventing some from taking part in population studies essential for tracking down disease-causing genes.

Some 90 per cent of America's six million Jews are Ashkenazi, a group descended from central and eastern European ancestors. Their intermarriage over the centuries has produced a population highly attractive to geneticists, for whom Ashkenazim, like Icelanders and Finns, provide a relatively homogenous group in which mutations are easier to detect.

But the results of early research have alarmed some Ashkenazim. Researchers have found that about 2.3 per cent of Ashkenazi women carry particular mutations predisposing them to breast and ovarian cancer; last September, the corresponding figure for colon cancer was published as 6.1 per cent.

Collins stressed that such findings do not indicate that Jews are any more predisposed to mutations than other populations, but merely that they have been the first to be studied. The "silver lining" for Ashkenazim, he added, is that they may be the first to benefit from drugs and other therapies that will derive from the work. **Meredith Wadman**



Flying fish: the oyster toadfish is part of the first Neurolab payload, launched two weeks ago.

Success prompts bid for second Neurolab launch

[WASHINGTON] US space officials hope next week to approve a second flight in September for the Neurolab space shuttle mission — whose first flight was successfully launched two weeks ago — to allow researchers to collect more data on how the nervous system adapts to spaceflight.

The US National Aeronautics and Space Administration (NASA) has already concluded that a reflight is "scientifically very worthwhile", says Edmond Reeves, head of flight planning for its Office of Life and Microgravity Sciences and Applications. A decision is expected by next Monday (4 May).

The international Neurolab mission, which includes participation by the National Institutes of Health and other agencies, is widely regarded as one of the best-designed space research missions ever flown.

A reflight would cost "tens of millions" of dollars, says Reeves, excluding the price of the shuttle launch. NASA and the other Neurolab participants — Europe, Japan and Canada — are already discussing how these costs should be split.

For its part, NASA would redirect funds intended for other microgravity and life-science research activities this year, as well as some funds from the shuttle account. To save time and money, the Spacelab module would be reflown with the same laboratory equipment and the same core crew of astronauts.

Only the scientist 'payload specialists' might change, Reeves says, as it could be useful to have different test subjects for some Neurolab investigations. It should also be possible to fly some experiments that were not included on the first flight.

The space research community is eager for as many 'gap filler' flights as possible while waiting for the space station to be completed in 2003. Another such flight is planned for October, but room for scientific experiments is limited. Commercial research payloads will take up 60 per cent of the US half of a small laboratory module

provided by Spacehab Inc. A second 'gap filler' mission in 2000 could fly a larger module, but so far there is no money for this.

A big uncertainty affecting the decision to reflly Neurolab is the space station construction schedule. NASA managers hope to learn during a joint programme review in Moscow this week whether a key Russian 'service module' — which keeps the station in orbit during the early assembly phase, and provides the living quarters for its first crew — will be ready for launch early next year.

Russian officials have said that the module, the third station component to be delivered, will miss its December launch by three or four months. But even that is uncertain.

An outside review team painted a gloomy picture of the station's future in a report last week. According to a task force set up by NASA's outside advisory council at the request of the agency's administrator Dan Goldin, the US investment alone will climb from \$17.4 billion to \$24.7 billion, and assembly will take between one and three years longer than NASA believes.

Space station managers continue to be too optimistic about schedule and finances, said the report, and NASA will need as much as \$250 million a year more than it has requested to keep the programme on track. "Budget and reserve levels have been, and continue to be, inadequate for a program of this size, complexity and development uncertainty, despite NASA's past contentions that the total funding level is adequate," the report said. The primary — but by no means the only — threat to the project is Russia's unreliability, according to the task force.

The report's conclusions were made public last month. But NASA's reaction has been muted. In Senate testimony last week, Goldin said only: "While we may differ on the level of criticality of specific issues raised, I believe the [task force] has captured important risk areas for NASA and the [space station] programme to consider." **Tony Reichhardt**

Troubled UK biotech firm faces new probe

[LONDON] The biotechnology company British Biotech is to be investigated for a second time by the London Stock Exchange following further revelations last week about its handling of data about sensitive clinical trials.

The investigation follows a decision by the US Securities and Exchange Commission to investigate the accuracy of statements about the company's main anticancer drug marimastat. Concern has also been reported about the company's decision to hold back news of difficulties in clinical trials with its other main drug, the pancreatitis therapy Zacutex.

These developments, as well as this month's sacking of Andrew Millar, director of clinical research (see *Nature* 392, 746: 1998), have seen British Biotech's share price nose-dive to less than £0.50 from a high of almost £3.00 in 1996, reducing the value of the company from nearly £2 billion to £330 million.

The company's shareholders are now waiting for British Biotech to break its silence. Failure to do so may strengthen calls for the removal of its founder and chief execu-

tive, Keith McCullagh. A statement from the company was expected this week.

The London stock market had previously cleared a move by three British Biotech directors, including McCullagh, to profit from the sale of shares in January 1995. The sale took place less than two weeks before the suspension of trials on batimastat, which was then the company's main anticancer drug. Batimastat was abandoned soon after.

It has now emerged that Millar circulated a memorandum as early as October 1994 calling for a halt to the clinical trials of batimastat following the discovery of side-effects.

Millar was dismissed two weeks ago for allegedly passing on confidential information to shareholders about the status of clinical trials with the company's latest flagship drugs. He has said that he was concerned that the company was not being open with shareholders about the progress of its drug trials.

The European Medicines Evaluation Agency (EMEA) wrote to British Biotech in May 1997 detailing its objections to an application for marketing approval for

Zacutex filed in February.

It said that apparently positive results from Zacutex trials in the United Kingdom did not provide conclusive evidence that the drug reduced deaths among sufferers of acute pancreatitis.

British Biotech issued a statement on the success of the UK trial result without alluding to EMEA's reservations. The stock market was not informed of EMEA's objections for another nine months.

Millar has said he was aware that positive results from UK clinical trials were not matched by interim results from a parallel — and larger — trial in the United States, and that the company's senior officials were made aware of this information as early as November 1996.

Shareholders, who include leading British fund management companies such as Perpetual and Mercury Asset Management, are expected to ask for a special meeting if the company's statement does not answer questions about the investigations, as well as the state of the company's flagship drugs and its overall strategy.

Ehsan Masood

Spanish government pledges full funding for Canaries telescope

[MUNICH] The Spanish government has pledged to fund fully a 10-metre optical telescope in the Canary Islands — the first large-scale research facility Spain has ever approved — even if no international partners are found.

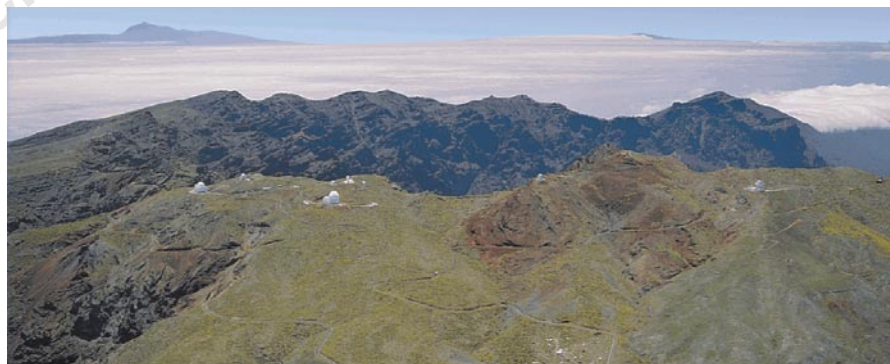
The decision puts an end to widespread fears in Spain that the government's refusal to make a commitment would delay the project for so long that it would no longer be an internationally competitive facility.

The telescope will be built on Roque de los Muchachos on the island of La Palma, one of two sites belonging to the Institute of Astrophysics of the Canaries (IAC). First light is planned for 2002, and the design will be based on the segmented-mirror structure of the Keck telescope in Hawaii.

After years of lobbying from the IAC, the government last year agreed to pay half the telescope's cost of US\$100 million, but only if the other half was first guaranteed by foreign partners.

Although the project had attracted interest from potential partners as distant as China and India, as well as from universities in Italy, Finland and the United States, none is thought to have felt sufficiently confident of the Spanish government's commitment to make their own funding promises.

Approving funding for the telescope was the first major decision by the newly created Office of Science and Technology (OST), which is in the office of the prime minister.



High hopes: the Roque de los Muchachos on La Palma where IAC's 10-m optical telescope will be built

"The office has given its definitive approval for the telescope," says Fernando Aldana, head of the OST. "That means that the national government, along with the regional government of the Canary Islands, will bear the whole cost of the project" if it fails to find international partners.

But he says that the IAC should continue to seek such partners to cover up to one-third of the project's costs, "given the scientific and technological contribution foreign experts will bring to it".

Francisco Sánchez, director of the IAC, says the government's declaration "will allow us to lift the brakes on negotiations" with potential partners. He is confident of being able to sign up partners before the end of the year, giving them time to provide input into the telescope's final design.

Sánchez says he is keener to bring partners on board for their expertise than for their money. Indeed, he says that there is more interest in the telescope from foreign researchers than the one-third maximum foreign participation the government wants.

For example, the University of Massachusetts is keen to use the telescope to follow up in the infrared spectral range the submillimetre observations made with the Large Millimeter Telescope 50-metre antenna that the university is building in collaboration with Mexico.

Viewing time offered to partners on the new telescope will be proportional to their financial contribution. Partners will also have access to the five per cent observing time on all 20 or so telescopes operating at the IAC's observatory sites.

Alison Abbott

Beijing media join attack on air pollution

[BEIJING] China's capital, one of the world's most polluted cities, has begun to clean up its air following moves by a number of cities in China to release data on air pollution levels. Beijing has long suffered from air pollution produced by the burning of low quality coal in homes and factories, and its growing number of cars is aggravating the problem.

Similar problems afflict many Chinese cities. Indeed, a World Bank report last year cited air pollution as one factor making chronic obstructive pulmonary disease — emphysema and chronic bronchitis — the leading cause of death in China.

The problems are not confined to China's cities. Acid rain is afflicting forests, particularly in the south. And China is also believed to be a growing source of the acid rain that falls in neighbouring countries, such as Japan (see *Nature* 392, 426; 1998).

Despite a 1989 law requiring environmental reports by all levels of the Chinese government, local governments have only recently begun to release figures on air pollution. Last year, 27 cities began making weekly air pollution reports, and Beijing's local government started releasing figures on 28 February, immediately before the opening of the National People's Congress in March.

The start of weekly reports in Beijing has been followed by intense media coverage of air pollution problems in what Western observers say is an orchestrated effort by the new central government to build public awareness as it starts to tackle the problem.

The former National Environmental Protection Agency has been upgraded to ministerial status in the central government, and renamed the State Environmental Protection Administration. But with only 400 employees it is hard for the new ministry to enforce environmental controls nationwide. Observers say this is why the administration is using the media to bring pressure on local governments.

The amount of information released varies from city to city. In Shanghai, one of

China's most progressive cities, the levels of three key pollutants — sulphur dioxide, nitrous oxides and total suspended particulates — are released each week. But in Beijing, only the level of the worst pollutant among the three is announced.

Last year's World Bank report said levels of sulphur dioxide and particulates in major Chinese cities such as Beijing, Shenyang and Xi'an were two to five times higher than the maximum recommended by the World Health Organization, making them among the most badly polluted cities in the world.

Among new activities in Beijing in recent weeks, the municipal environmental protection and traffic administration bureaus have begun spot checks of cars to find out if they

have catalytic converters, fining drivers of the worst polluting cars.

The local Beijing authorities have also announced that they will use 1.5 billion cubic metres of natural gas over the next two years in a bid to discourage residents from burning coal bricks in home stoves. The city plans to set up 40 coal-free zones and encourage the use of high grade coal elsewhere, as well as introducing central heating to 50 per cent of homes by 2000.

The burning of coal in homes is one of main sources of air pollution. The 27 million tons of coal that Beijing burns each year produces a haze with high levels of sulphur dioxide and dust that hangs over the city and chokes residents.

David Swinbanks

Banned drug 'still used until this month'

[WASHINGTON] The New York research institute recently criticized for using a now-banned diet drug in experiments on young children has revealed that until this month the drug, fenfluramine, was still used in separate psychiatric tests with adolescents.

The study that first drew attention, non-therapeutic research into aggression in poor, young, non-white boys, was actually stopped in 1995, two years before fenfluramine — part of the diet-drug combination Fen-Phen — was withdrawn from the market after evidence that it caused serious heart-valve damage in adults who used it for months (see *Nature* 392, 747; 1998).

But the New York State Psychiatric Institute continued a separate study using the drug until this month, with approval from the US Food and Drug Administration (FDA). Claudia Bial, a spokeswoman for the Manhattan-based institute, says that the goal of this work was "to determine if adolescents with major depressive disorder, and who attempt suicide, show evidence of abnormal brain-serotonin function".

Fenfluramine increases brain serotonin levels. Bial says that the study, funded by the National Institute of Mental Health, was begun in 1992, and had enrolled 98 subjects before it ceased temporarily when the FDA withdrew fenfluramine in September, 1997.

After researchers gained FDA permission, the study was restarted in March this year, and two more subjects received single doses of fenfluramine before the investigators stopped the study this month because of the storm unleashed by local press reports of the work on children.

The FDA defended both studies at a congressional hearing last week. While declining to comment on the scientific significance of the work, Michael Friedman, the acting FDA commissioner, said he felt it

was "appropriate and ethical and that it is being conducted under carefully scrutinized circumstances". But, he added, "I can assure you that I'm going to be looking into it much more carefully."

Friedman told the House of Representatives Government Reform and Oversight Committee that after fenfluramine was recalled from clinical use as a diet drug, researchers using it were allowed to continue on three conditions: that informed consent forms were revised to reflect the risk of heart damage (parents sign these on behalf of their children); that echocardiograms of subjects' heart valves were conducted before and after the research; and that local institutional review boards consented to the research in the light of the new findings about valve damage.

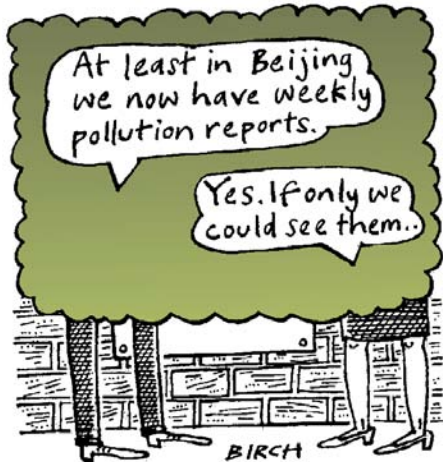
Bial said these conditions were met, and provided an informed consent form for the adolescent study that warned of "potentially life-threatening ... heart-valve damage". She said that echocardiograms on the two recent subjects revealed no damage. Friedman also stressed participation in the research was "voluntary". He told committee chairman Dan Burton (Republican, Indiana) it was "a question of personal choice".

But Burton said that he could not understand the government's approval of a child's exposure to a drug whose effects in adults had proved so damaging.

Friedman said that there had "never" been a case of heart-valve damage from one dose of fenfluramine in an adult, but conceded that data on children was scant.

Adil Shamoo, of Citizens for Responsible Care in Psychiatry and Research, an advocacy group, told the hearing that the researchers had "no therapeutic or medical cause" to use fenfluramine on the youngsters.

Meredith Wadman



Unesco board set to agree compromise on bioethics committee

[PARIS] Uncertainty over the future of the International Bioethics Committee (IBC) of the United Nations Educational Scientific and Cultural Organization (Unesco) is likely to be resolved at a meeting of the UN agency's 58-member executive board this week.

The IBC was set up in 1994 with the task of drafting the first UN text governing science and human rights, the "universal declaration on the human genome and human rights". But that remit ended with the adoption of the declaration last November by the 186 member states of Unesco.

At the same time, the declaration gave the IBC the new role of overseeing the implementation of the text's provisions, a set of broad principles affirming the need to protect the individual against genetic discrimination while upholding the principle of scientific freedom.

Member states have now rejected a Unesco proposal that the IBC, which is made up of independent experts appointed by Unesco director-general Federico Mayor, be replaced with a committee of both experts and civil servants from member states.

Under a compromise to be submitted to the executive board, the 36-member IBC will remain independent, while a separate committee representing member states will be set up in parallel. But how this arrangement will work in practice is far from obvious.

One new activity for the IBC, for example, will be to release 'opinions' on ethical issues. But it is not clear whether the IBC will be able to make these public immediately, or whether they will first have to be approved by the second committee.

In negotiations this month, the compromise proposal won broad agreement, with support, for example, from France, the United Kingdom, Japan and Israel — as well as the United States, which is not a member of Unesco. Germany favoured a political committee, however, apparently on the grounds that this would allow greater control.

Meanwhile, Unesco's executive committee will also be asked to approve a proposal to create a separate World Commission on the Ethics of Scientific Knowledge and Technology, which would consider ethical issues in areas outside biomedicine. The commission would be chaired by Vigdis Finnbogadóttir, a former president of Iceland.

Under preliminary proposals, the commission's goal would be to promote debate among scientists, intellectuals, public and private decision-makers, and the public on the ethical and risk issues in such areas as energy, the use of freshwater resources and the information society.

Declan Butler

Free market is advocated for Australian universities

[CANBERRA] Radical recommendations in the final report of a review of financing and policy for higher education in Australia have set the scene for a heated debate on university policy in the run-up to the next general election, due within 11 months.

A seven-member panel chaired by Roderick West, a retired head teacher of a private school, is advocating opening the 37 universities completely to market forces through deregulation and tuition fees.

The report contains no financial analysis of the impact of its proposals — an omission that has drawn widespread criticism. But there is stronger coverage of research issues, given little attention in a draft released last November (see *Nature* 390, 212; 1997).

Teaching and research are more specifically linked, and the final report stresses the need for "strategic planning, priority setting and greater coordination of national research effort", in contrast to the 'market' philosophy it applies elsewhere.

The main recommendation for students is that they should receive 'entitlements' from the government at standard levels set for three broad grades of courses (basic, laboratory-based, clinically based) which they could 'cash' at the university of their choice. Universities would compete for enrolments, and widely differing fees are expected.

This would replace the system under which universities receive grants according to government-approved numbers of students, which are supplemented by partial fees paid by students after they graduate.

In addition, universities could capitalize their assets (which would be audited) and then sell or borrow against them. But they would be liable for taxation like businesses.

Asked how a deregulated system would address the crisis said to be facing science in universities, the West committee said that science enrolments have been increasing and expressed confidence that "the market" would sort out any short-term problems.

Vice-chancellors reacted cautiously, but John White, policy secretary for the Australia Academy of Science, says that, although West "perceptively diagnoses serious structural and financial problems within the system", the remedy "is narrow in its philosophical base and even paradoxical". Academic and student bodies criticized the report as "regressive and irresponsible".

The committee says its goal is to turn Australia into "a lifelong learning society" through a new financing and regulatory framework to cope with increased expectations for higher education. It is confident that information technology will "revolu-

tionize" management and what it calls the "education products" of universities.

Speaking at the report launch on 17 April, West declared faith in the ability of students to select courses and universities, and to have "a real say in what universities provide". If adopted, he said, the committee's scheme would "avoid negotiation between govern-



West: wants students to drive funding.

ment and universities as students drive public funding and Canberra bureaucrats are taken out of the equation".

Noting that "current funding levels are causing tension", the review's scheme is based on no overall increase in public support. But it still believes government should maintain, over five years, the average of public funding per equivalent full-time undergraduate student.

In a rare quoting of figures, the committee says that, on current trends, the level of overhead support for project grants will decline from 28 cents in the dollar (the US figure exceeds 50 cents) to 12 cents by 2000, and could only be maintained with an additional A\$13 million (US\$8.5 million) in 1999 and A\$55 million a year from 2000. It says an increase is "warranted".

The most significant reform in research would see the Australian Research Council established as "an independent body" with "a wider range of functions and increased accountability and transparency". Currently part of the Education Department, the ARC would be responsible for setting national university research policy after being given "explicit authority to determine priorities within its own programmes".

But the review's advocacy of nationally directed priorities in funding for university research contrasts with support for a "student centred" approach to funding research training, where places would be competitive.

The review's prospects of gaining government support for its key proposals have been placed in serious doubt by the education minister, David Kemp, who repeated his earlier rejection of 'vouchers' for students and 'deregulated fees' for universities.

Kemp's first major speech on universities has declared his own agenda for reform, indicating that he is likely to take note only of those of West's recommendations that fit his preferences, such as linking student choice with funding mechanisms.

Peter Pockley

PETER POCKLEY

Test ban treaty faces make or break in the US Senate

Colin Macilwain

A tough time lies ahead for the US administration in its push to achieve the two-thirds Senate majority needed to ratify the treaty banning atomic weapons testing. Republican senators are digging in with a complicated set of conditions and will prove hard to shift.

[WASHINGTON] The fate of the Comprehensive Test Ban Treaty (CTBT) at the hands of the US Senate is likely to be decided within the next month, as the Clinton administration begins its bid to overcome strong Republican opposition and get the treaty considered and ratified.

Senior administration officials—including President Bill Clinton himself and the Secretary of State, Madeleine Albright—are expected to begin their long-awaited drive for ratification as soon as the Senate completes its deliberations on the expansion of the North Atlantic Treaty Organisation (NATO). This is likely to happen within days.

But many Senate staff on both sides believe that, even if the debate takes place, the treaty will not be ratified this year. “I don’t give it much of a chance,” says one aide to the Senate Democratic leadership. “A handful of people are able to hold this up, and we haven’t been able to devise a way to put pressure on them.”

The treaty to ban atomic weapons testing was approved by the United Nations General Assembly in 1996, but will not take effect until it has been ratified by 44 specified signatories; it requires a two-thirds majority in the US Senate. It has been the goal of many nuclear weapons scientists for four decades,

and is strongly supported by several scientific organizations.

Advocates argue that the administration is in a good position to force—and win—a public debate on the treaty, and to get Republican senators to bring it forward to a vote. They are confident of obtaining the necessary two-thirds majority in such

a vote, because of strong public sentiment against nuclear testing.

But political opponents of the treaty, including in particular Senator Jesse Helms (Republican, North Carolina), chair of the powerful Foreign Relations Committee, are



Will he play ball? Secretary of State Madeleine Albright has to win round Republicans such as Senator Jesse Helms if test ban is to progress.

well entrenched, and a complex set of conditions must be met if the treaty is to be ratified.

At least five Senate committees have an interest in the treaty. But three of the most powerful—the Armed Services, Intelligence and Foreign Relations committees—have yet to hold hearings on it.

In January, Helms informed Clinton that his committee would not consider the CTBT until it had been given an opportunity to review two other foreign policy items, the Kyoto protocol on greenhouse-gas emissions and a set of proposed modifications to the Anti-Ballistic Missile (ABM) treaty.

Clinton responded in February, saying that there would be no movement on Kyoto, but implicitly raising the prospect that the ABM protocols may be on offer as part of some sort of deal. Clinton said he wanted the CTBT ratified in time for his planned visit to the Indian subcontinent later this year.

According to lobbyists for the treaty such as Daryl Kimball of the Coalition to Reduce Nuclear Dangers, the prospects for ratification will improve if Russia’s Duma, the lower house of its parliament, ratifies the START 2 treaty limiting the number of nuclear weapons held by Russia and the United States. It may do this in the next two months.

But the Duma’s ratification of START 2 will probably be conditional on the retention of the ABM treaty in its existing form. Republicans in the Senate are keen to modify the ABM treaty in ways that would offend Russia—for example, by allowing deployment of ballistic missile defence in ‘small theatre’ conflicts.

Helms has given no hint of what will shift him. Lobbyists for the treaty say that any progress is likely to depend on a deal between Clinton and Trent Lott (Republican, Mississippi), the leader of the Senate Republicans, which might force Helms to modify his position.

Last year, in analogous circumstances, an agreement between Clinton and Lott effectively forced Helms to release the Chemical Weapons Convention, which was duly ratified by the Senate. But sources close to Helms say that whereas Republicans were split on the convention—which had been negotiated by President George Bush—they are relatively united in their opposition to the CTBT. The only prominent Republican senator who has publicly supported it is Pete Domenici (Republican, New Mexico).

The key question for treaty advocates is whether the administration’s pending publicity offensive will raise the profile of the treaty sufficiently to fracture that unity. Spurgeon Keeny, for example, president of the Arms Control Association, says the administration cannot lose by vigorously pursuing the treaty.

“If they get it, it is an important accomplishment for the administration; if the Republicans fight it, it is good politics,” says Keeny. “A lot of Republicans don’t want to be put on the spot on this issue. But I don’t think they are going to have that luxury, because the administration is going to stick it to them.”

Officials at the White House National Security Council decline to comment on their strategy. Clinton sent the treaty to the Senate last September, and announced in his January State of the Union address that he wanted it ratified during 1998. The chances of that are probably better than they will be in 1999, when the Republicans are likely to have a larger majority, and the president may be weakened by the approaching end of his tenure.

Only countries that have ratified the treaty will be allowed to take part in a conference in late 1999 which will decide the future of the treaty if, as expected, India, Pakistan and North Korea have failed to sign it by then. □



Domenici: lone voice among Republicans.

Oil industry lobby plans rival to UN climate science panel

[WASHINGTON] Members of the anti-global-warming lobby in the United States propose an industry-funded rival to the United Nations panel of climate scientists. The rival panel would seek to convince the public that the Kyoto agreement to reduce greenhouse gas emissions is based on bad science.

The proposal comes from an unofficial group of public relations officials from oil companies Exxon and Chevron and the American Petroleum Institute. They suggest that \$5 million should be spent recruiting 20 scientists to set up a Global Climate Science Data Center as "a sound scientific alternative" to the UN-backed Intergovernmental Panel on Climate Change.

Of that sum, \$600,000 would be spent on a public affairs offensive in the run-up to this November's meeting of the parties to Kyoto in Buenos Aires. Five scientific 'new faces' would be media-trained to attack the theory of man-made climate change.

The initiative, details of which were leaked by an environmentalist group, coincides with a decision by the British/Dutch oil company Shell to leave the Global Climate Coalition, the main US lobby group that opposes the climate treaty. Shell's decision follows that of BP, announced two years ago (see *Nature* 383, 470; 1996).

Chemical industry boost for German research

[MUNICH] Germany's chemical industry last year increased the number of scientists it employed in research and development for the first time since 1991, with the total rising by 1.4 per cent. The number of young scientists in the sector also rose in 1997, with an additional 1,900 employed — 9.5 per cent more than in 1996, according to the Association of the Chemical Industry (VCI).

The industry's spending on research carried out at home increased by DM400 million (US\$220 million) to a total of DM11.7 billion, or 6.2 per cent of turnover (a further DM5 billion was spent on basic research abroad). The trend of boosting research in Germany is the result mainly of reduced bureaucracy in laws on emissions and gene technology, says the VCI.

Science likely to lose its autonomy in Russia

[MOSCOW] Russia's Ministry of Science and Technology is likely to lose its autonomous status, according to unofficial reports from Moscow. The ministry is expected to be linked either with the Ministry of Education or made a department in the Ministry of

Economy. Both ideas are understood to be part of President Boris Yeltsin's drive to reduce the state's civil-service apparatus.

There are two candidates to head Russia's science programme. One is Mikhail Kirpichnikov, head of the department of science and education in the present cabinet and recently elected to the Russian Academy of Sciences. The other is Ivan Bortnik, director of the foundation for promoting small enterprises in science and technology, and a former deputy science minister.

France diverts funds to start-up innovators

[PARIS] Innovation in France is set to receive a substantial boost with a decision to invest a proportion of the funds from a popular endowment scheme in the country's stock market for start-up companies. Finance minister Dominique Strauss-Kahn has announced that five per cent of the funds of Assurance Vie will be invested in the Nouveau Marché — the French equivalent of NASDAQ in the United States.

The announcement comes in the run-up to a government-organized national colloquium on innovation, to be held on 12 May, which will be attended by Prime Minister Lionel Jospin. The government is expected to announce a move that has been demanded by many scientists — the end of laws forbidding publicly funded researchers from holding shares or sitting on the boards of companies with which they have research links (see *Nature* 392, 214; 1998). This restriction is widely viewed as an obstacle to entrepreneurial culture in France.

Cambridge wins money for research centres

[LONDON] The University of Cambridge in the United Kingdom has secured £32 million (US\$53 million) from the chemical and food company Unilever and British Petroleum (BP) to fund two professorships and two new research centres.

Unilever will contribute £13 million over five years towards a centre for research into molecular sciences which, among other functions, will provide worldwide access to a databank of discoveries. BP is to give £1.3 million a year for 15 years to set up the BP Institute, to investigate the flow of oil, gas and water in the Earth.

The sponsorships surpass last year's £12 million donation by Microsoft chief Bill Gates for an advanced computer laboratory.

Antibiotics resistance 'is major threat to health'

[LONDON] Resistance to antibiotics has become a "major threat to public health" but little funding is available for research into

antimicrobial resistance, according to a report from the science and technology committee of the United Kingdom's House of Lords.

The report says the UK Medical Research Council (MRC) spent only £300,000 on antibiotic resistance during 1995–96, whereas in the same period £60 million was spent on research into immunology, infections and inflammatory disease.

The MRC told the Lords committee that there was a lack of high-quality research proposals. But Tessa Jowell, minister for public health, said that this was not necessarily the case. "The way in which research budgets are presently constructed does not necessarily mean that public health issues are given a proper opportunity to bid for resources," she said.

Japan set to support asteroid spotter

[TOKYO] The Japanese government's announcement last week that it would support the construction of an optical telescope and radar for detecting threatening space debris and near-Earth objects (NEOs) has been welcomed by space scientists as a contribution to international 'spaceguard' activities. The government had been lobbied to build such a telescope by the Japan Spaceguard Association.

The telescope, with an aperture of 1 metre, will detect asteroids that could collide with the Earth, as well as space debris such as spent boosters. The radar will be able to detect smaller debris in low Earth orbit at 400–1000 km, the same altitude as the planned international space station. The total project, expected to cost ¥2 billion (US\$15 million) over five years, will be funded by the Science and Technology Agency (STA).

Control engineer takes control in Australia

[CANBERRA] An engineer has for the first time become president of the Australian Academy of Science. Brian Anderson today (30 April) succeeds Sir Gustav Nossal for a four-year term. Anderson is the founding director of the Research School of Information Sciences and Engineering at the Australian National University in Canberra.

The 57-year-old Anderson is one of the most productive research engineers in Australia, with seven books and about 500 papers to his name, and specialises in electrical networks, communication and control systems and signal processing. He admits to concern about the limited support for research in Australia. "Until you can repair that underfunding, it is a bit illusory to talk about research policy," he says.

Tycho's illusion and human cognition

Sir — For several decades at the start of the seventeenth century, the main cosmological alternative to the Sun-centred Copernican system was not the Ptolemaic, but a compromise system proposed by Tycho Brahe (Fig. 1). In the Tychonic system, the planets move around the Sun exactly as Copernicus proposed. But the Sun, accompanied by the orbits of the planets, moves around the Earth. This new system appeared to embrace the elegance of Copernicus yet avoid his heresies. But in this system the orbit of the Sun and the orbit of Mars intersect, so it seemed obvious both in Tycho's time and our own that the system is inconsistent with the traditional view that the Sun and planets are carried by solid spheres.

In fact, the collision seen in the Tychonic diagram is an illusion, which has deceived the experts for 400 years. From Kepler in Tycho's time to Thomas Kuhn and many others in our own, there has been no dissent. Why?

Figure 1 separates Tycho's original diagram into two components: the cut-out, Fig. 1a, contains everything that shares in the Sun's annual motion and nothing else; Fig. 1b is what is left of Tycho's diagram after removing the cut-out. To run the model, make an enlarged photocopy of the figures, cut the two apart, and then fold the cut-out to bisect the black circular region in its centre. Next, cut the black centre out and trim away the remaining black on the periphery. Then unfold the cut-out and centre its hole over the Earth–Moon region of the template. Turning the cut-out will produce the Tychonic orbit of the Sun, accompanied (as Tycho and observations require) by the orbits of the planets.

On a scale of about 1 to 10^{12} , the cut-out is a model of a single Tychonic sphere that carries six epicycles (the five planetary orbits plus a solar epicycle too small to show). This replaces Ptolemy's six spheres,

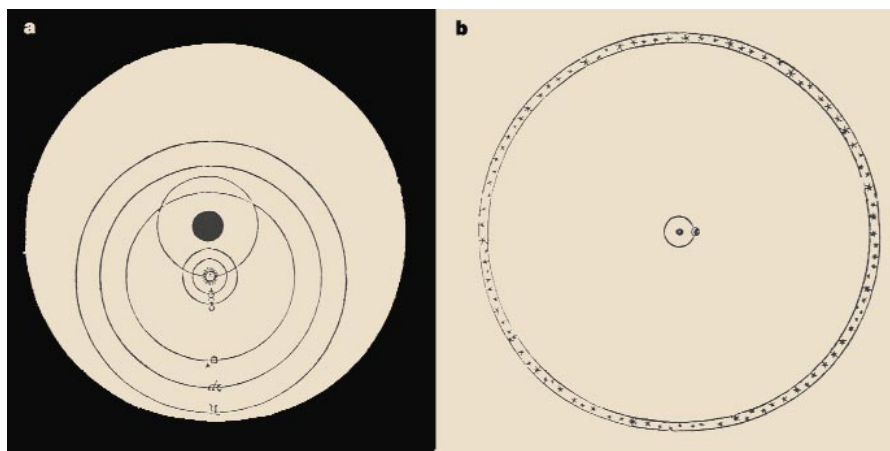


Figure 1 The Tychonic sphere in two parts. See text for instructions.

each carrying one epicycle. The single Tychonic sphere carries both the Sun and its tributary orbits in the common annual motion that the logic of the system requires. Within that single sphere, the heliocentric orbits would be physical objects carrying the planets. But the path of the Sun in Tycho's diagram does not carry anything — Mars could no more collide with that mathematical locus than a ship could collide with the equator. But an expert in this matter, even more than a non-expert, usually cannot believe that until actually seeing the cut-out move. Then the illusion of an unavoidable collision is destroyed, and it becomes hard to believe that for 400 years no one questioned the collision.

What makes the illusion? Our experience is overwhelmingly that little things circle big things. But with Tycho's system the large orbits of the outer planets must swing around the Earth on the smaller radius shared with the Sun. If a person tries to imagine that rotation, intuition about how things move rebels. In the mind's eye, the Sun goes around the Earth accompanied by the smaller orbits of Venus and Mercury (as it should), but the larger

orbits stay put, which they should not. This discrepancy cannot be conscious, for a person would then recognize it as logically absurd. But with the cut-out, it is easy to see that the orbits are carried by rotation of the entire region between the Earth/Moon system and the fixed stars. Now we see big carrying small, which feels intuitively comfortable, and the illusion disappears.

There has been much debate among psychologists about whether cognitive illusions can provide real insight into human reasoning, or reveal only how clever experimenters can trick naive subjects. But Tycho's illusion is plainly no mere laboratory trick. It has been deluding the experts for 400 years, and on a question they have taken very seriously.

A full paper discussing the illusion and its cognitive implications is available at <http://www.harrisschool.uchicago.edu/Tycho.html>.

Howard Margolis

Harris School,
University of Chicago,
1155 E 60th Street,
Chicago, Illinois 60637, USA
e-mail: hmarg@uchicago.edu

Art in the round

Sir — How lovely to see the word “bollocks” appearing, perhaps for the first time, in *Nature* (392, 663; 1998). And how appropriate that it should occur in Martin Kemp's series of pieces exploring the assimilation of scientific ideas into artistic creation.

It has been said that scientists tend to have a greater appreciation of art than artists have of science. Kemp's illustrations, together with his report of the unnamed biology graduate student's thoughtful appraisal of Cornelia Parker's work,

provide a powerful counter-argument to this assertion.

A fellow-scientist of my acquaintance recently ridiculed a monochrome painting at an exhibition into which she had presumably stumbled by mistake, or to escape from bad weather; superficially ‘simple’ modern works — Mark Rothko's canvases and Carl Andre's sculptures come to mind — are especially vulnerable. I suggested to my colleague that an article in *Nature* might be no more intelligible to the painter than this painting was to her, but that the artist would be unlikely to dismiss the article as casually

as she had the painting.

If the University of Leicester student is working towards a PhD, and if it is not too late, I would urge the student's supervisor to invite Cornelia Parker to act as an external examiner for the thesis. Ms Parker can hardly be any less eligible to evaluate this student's work than the student is to assess hers.

She could announce her verdict in the departmental tea room.

Ian Smith

18 Lockwood Street,
York YO31 7QY, UK
e-mail: i.smith@csl.gov.uk

British forensic science in the dock

Unregulated forensic science practices have led to a spate of wrongful convictions. There are too many 'cowboy' practitioners whose services can be bought at a price.

Zakaria Erzinçlioglu

In the United Kingdom, miscarriages of justice resulting from faulty or incompetent forensic science practices began to attract public notice during the 1980s. Cases such as the 'Birmingham Six', the 'Maguire Seven', Judith Ward, Stefan Kiszko and John Preece became notorious. The widespread political and public concern resulted in several inquiries into the state of forensic science. These included the Royal Commission on Criminal Justice¹ and the Report on Forensic Science by the House of Lords Select Committee on Science and Technology². The general conclusions of these inquiries were that, although there was room for further improvement and reform, the forensic science profession had set its house in order.

Soon after these reports were published, the discovery that contaminated centrifuges at the Royal Armament Research and Development Establishment had been used to secure convictions in 12 terrorist cases raised the question of whether further miscarriages of justice had taken place. Once again, the reliability of forensic science came under scrutiny. Nevertheless, it was generally agreed that this was an isolated case and that British forensic science was now in a healthy state.

I cannot agree with this view. Forensic science in the United Kingdom is in a very poor and disordered state. There are two separate, though related, problems. First, standards — both scientific and ethical — have fallen and will continue to do so unless the present structure and organization of the profession are reformed. Second, the way forensic evidence is handled after leaving the laboratory often results in its perversion, through handling not only by the police but also by lawyers in court.

Enter the marketplace

How did the fall in standards come about? Until 1989, the Forensic Science Service (FSS) was an integral part of the police department at the Home Office. Police forces routinely submitted their cases to FSS laboratories, where investigations were done without payment of a fee, although the police forces made an annual contribution to the service. This fixed contribution was made on the basis of the size of the force, not on the amount of forensic work carried out for the force in question. But in 1989 the House of Commons Home Affairs Committee recommended that the FSS be given agency status; in effect, it was privatized. Police forces have since had to pay the full



cost of forensic work on a case-by-case basis. The FSS is now funded entirely by case work carried out for the police and others.

The FSS must now compete with other practitioners in the marketplace, many of whom are either incompetent or dishonest. Forensic science is totally unregulated and sharks and cowboys abound. Apart from the specialist field of forensic medicine (or pathology), there is nothing to prevent anyone — qualified or not — from advertising themselves as a forensic scientist, whether purporting to be a specialist in toxicology, ballistics, DNA, document forgery, blood analysis or whatever.

It may be argued that unqualified and inexperienced people would not survive in the marketplace, because everybody would see them for what they are and not consult them. This presupposes that those who consult forensic scientists are always interested in the truth, which is far from being the case. Cowboy consultants exist largely to muddy the waters in court; there is a market for their services. An unscrupulous lawyer, wishing to bolster his client's case with what is meant to

be seen as 'respectable' scientific evidence, may consult such quack practitioners if it helps them to win cases and earn more money.

Unscrupulous practices

I have come across innumerable cases in which fraudulent practitioners have confused matters in court to such an extent that the jury was unable to evaluate the true significance of the evidence. The inevitable result is that the evidence is dismissed altogether. In the worst cases, the evidence of a charismatic cowboy consultant will be believed. In neither situation is justice served.

Many practitioners are completely out of their depth. I have known several who are unfamiliar with basic scientific principles, presenting conclusions that are physical impossibilities. Some consultants are proud to boast that they are "uninhibited by the truth", as one put it. A practitioner at a recent trial invented new concepts for which there is no basis in fact. Such individuals make a very comfortable living indeed.

The existence of these practitioners is largely a function of the adversarial system, in which two sides argue, or 'fight', their cases in court. When one side engages a specialist, whose report is then made available to the other side, the 'opposition' then seeks the advice of another specialist who is willing to contradict the conclusions of the first specialist for a fee.

Moreover, the way in which scientific evidence is treated in court sometimes results in the jury being misled. I have seen science, or what is presented as science, being used by

Standards — both scientific and ethical — have fallen and will continue to do so unless the structure and organization of the profession are reformed

barristers to win cases regardless of the truth. On one occasion I was ferociously cross-examined by a barrister who came up to me when the day's proceedings were over and said that he enjoyed listening to my evidence and would I please send him copies of my publications in forensic science! It was clear that he accepted my conclusions completely, but it did not serve his purpose to admit it in court.

At a recent murder trial the consultant retained by the defence asserted that he could determine the exact time of death of the victim by using a particular technique. Having myself researched this technique and applied it in case investigations for more than 20 years, I know that this is quite impossible. Only a minimum time of death can be established, that is, a time by which death must have occurred. I explained the man's errors to the barristers, so that no doubt could possibly have remained in their minds, but all my explanations were ignored.

All these ills flow from the marketplace approach to forensic science, yet those who support this state of affairs assert that the best person to evaluate a 'product' is the 'customer'. Forensic science advice (the product) is best evaluated by the lawyer or police officer who pays for it (the customer), or so the argument goes. If one buys a pound of apples from a grocer and they turn out to be rotten, one is entitled to go to another grocer whose apples are fresh and edible. So is it not reasonable that forensic science be evaluated in the same way?

No, I do not think it is reasonable. The provision of apples and the provision of forensic science advice are not comparable. The consequences of supplying bad apples affect the customer alone, whereas the consequences of bad forensic science affect most closely a third party, the defendant at a trial, who had nothing to do with the choice of the scientist in question. Satisfying a solicitor does not necessarily mean that a competent forensic investigation has been carried out; it means only that the solicitor was given what he wanted. This usually means ammunition with which to win his case, regardless of whether the work done was competent or honest. A forensic scientist who is prepared to please his customers in this way will be consulted again and again. The notion that forensic science is a product to be evaluated like any other is quite incomprehensible.

The problem is more complex than this, however, for it is not only the openly discreditable practitioner who represents a threat to justice. There is great financial pressure on any forensic practitioner working within the adversarial system. Even honest, competent scientists have to earn a living, yet the pressure on them to produce results pleasing to those who consult them means there is a danger of their occasionally compromising their principles. The pressure on forensic

scientists from lawyers and the police can be very strong.

What did the various reports recommend? The House of Lords select committee made two important recommendations. First, it proposed "a system of individual registration of all forensic scientists. Scientists should be registered according to speciality, and at one of two levels. Anyone should continue to be allowed to practice, but it would be an offence to purport to be registered when not, and expert evidence from unregistered persons should become exceptional. Registration should depend on qualifications, experience, references and a casebook; it should be subject to review and withdrawal. It should be administered by the Government, with the help of a small Board, and delegated to the appropriate professional body wherever possible." The government undertook to consider this recommendation, but nothing was done.

The second recommendation of the select committee was that there should be established a "new research centre, centrally funded to conduct research for the benefit of the whole forensic science community.... In the event of further proceedings along the lines of the recent consideration by the Court of Appeal of the cases of the Birmingham six, the Maguire seven and Miss Judith Ward, such a centre might provide Ministers and the courts with a source of scientific advice which could be seen to be wholly independent."

Two British universities attempted to set up such a centre, but both attempts failed, mainly because the government would not support them financially. Regrettably, the universities in question also felt unable to back their initiatives with the necessary funds.

The Royal Commission recommended the establishment of a Forensic Science Advisory Council to monitor standards. In its response¹, the government said that it "did not see the need" for such a council. Finally, Brian Caddy of Strathclyde University produced a report² on the contaminated centrifuges at the Royal Armament Research and Development Establishment (now the Forensic Explosives Laboratory); in its response, the government said it would consider establishing a Forensic Science Inspectorate. Nothing has yet been done.

Earlier this year, the report of the Forensic Science Working Group recommended the establishment of a Registration Council that would maintain a register of forensic scientists who have been assessed on their competence. Full proposals on the funding and organization of this council were presented by the working group.

I welcome this proposal, but with reservations. Although the proposed council and register would weed out most incompetent practitioners, it will still be faced with the problem of assessing those who are dishonest, because it is much easier to assess some-

one's competence than their integrity. A practitioner may be competent yet dishonest; it would be very difficult to exclude such people from the register without a messy exposure should they apply to be registered.

Restoring justice

So what can be done? Two things seem to me to be necessary. First, the use of science as a barrister's tool should be brought to an end, because forcing forensic scientists to support one side against the other is one of the reasons why fraudulent practitioners exist. Many lawyers want to carry on using them, but I do not see why the forensic witness should be answerable to one side or the other in court. The forensic scientist should be answerable to the court alone. After all, the judge himself is a neutral participant in the adversarial process and there is no reason why forensic scientists should not also be neutral. Of course, when I say 'answerable to' the court, I mean 'paid by' the court. He who pays the piper calls the tune: a barrister's call for a tune is not disinterested, but the judge's call is. If this idea is implemented, the market for quack practitioners, as well as the financial pressure on other forensic scientists to compromise themselves, will disappear.

The second requirement is a fully staffed, statutory body of forensic science that would be answerable solely to the judiciary. Although police forces and lawyers would go to it for advice, the body would not depend on the goodwill of its 'customers'. The easiest way to implement this proposal would be to reverse the agency status of the FSS and place it under the control of the judiciary.

Despite the abundance of malpractice within British forensic science, there is still a great deal of good that can be harnessed to the service of justice. The FSS is an excellent institution, and several other government and independent laboratories also have high standards. It has also been widely used as a model by other countries. But regulation of forensic science is more stringent in the United States, France and Germany, countries that could themselves now act as models for the British system in this respect. The problems within the profession could, in principle, be solved very easily. The main difficulty lies in convincing those who need to be convinced that a serious problem exists at all. □

Zakaria Erzinçlioglu is at 28 Harlton Road, Little Eversden, Cambridge CB3 7HB, UK. He practised forensic science for more than 20 years, and was a senior research associate at the University of Cambridge.

1. Report of the Royal Commission on Criminal Justice. Cm 2263. (1993).

2. Forensic Science Report. House of Lords Select Committee on Science & Technology. House of Lords Session 1992-93 5th Report. HL Paper 24 (1993).

3. Final Government Response. Hansard col 185 (1997).

4. Caddy, B. Assessment and Implications of Centrifuge Contamination in the Trace Explosive Section of the Forensic Explosives Laboratory at Fort Halstead. CM 3491 (1996).

A marriage of bone and nacre

Peter Westbroek and Frédéric Marin

Clinical experiments aimed at reducing bone loss in jaws raise a raft of interdisciplinary issues. Those investigating biomineralization, early evolutionary history and biological signal transduction might take an interest.

Our immune system is like a zealous gardener who spots and wipes out all that does not belong in his garden, even the tiniest of weeds. A group including Gérard Atlan (Montpellier) and Evelyn Lopez (Paris) have conducted an experiment the results of which are astonishing and show utter disregard of this system. Since their publication last year¹, however, their significance has passed largely unnoticed.

The authors took mother-of-pearl (nacre) from the bivalve mollusc *Pinctada maxima*, ground it to a fine powder, and mixed it into a slurry with the blood of eight female patients, between 48 and 55 years old, who were suffering from bone loss in their upper jaw. The slurry was then injected into tissues where the bone was missing. Biopsies taken after six months showed no evidence of inflammatory reaction; the nacre had been accepted by the tissues as if it had been taken from the patients themselves (Fig. 1). Even more surprisingly, bone-forming cells (osteoblasts) had been activated and new, healthy bone had formed throughout the implanted material, while the nacre particles were slowly dissolving away. The fresh bone tissue was tightly welded to the nacre particles without any intervening soft or fibrous tissue. Osteoclasts, cells that dissolve bone, were active around the bone, but did not touch the nacre. Moreover, these results have been confirmed independently using nacre and rat bones². So here, on the face of it, we have an implantation material of unparalleled virtues.

What has nacre in common with bone and teeth to be able to fool our immune system? At first sight, the differences seem greater than the similarities. Nacre is pure calcium carbonate, of the aragonite form, whereas bone and teeth consist mainly of hydroxyapatite (calcium phosphate). On the other hand, both biominerals are products of animal tissues and in both cases a complex macromolecular matrix is closely associated with the constituent mineral crystallites.

In earlier work, Lopez and her collaborators methodically followed up the hypothesis that implanted nacre issues chemical signals from its matrix that can activate osteoblasts. They first placed small fragments of nacre on top of a layer of osteoblasts

cultivated in a Petri dish with no other activators³. The osteoblasts started to proliferate and were attracted by, and subsequently attached to, the nacre chips. Some bone-like mineral was deposited at the interface. This recognition of nacre by the osteoblasts is notable, as synthetic aragonite did not have a comparable effect.

In an extension of this experiment, chips of nacre and bone were placed 1 mm apart on the layer of human osteoblasts⁴. The fragments started to grow out towards each other, and after four to six weeks produced a dense tissue that filled the intervening space.

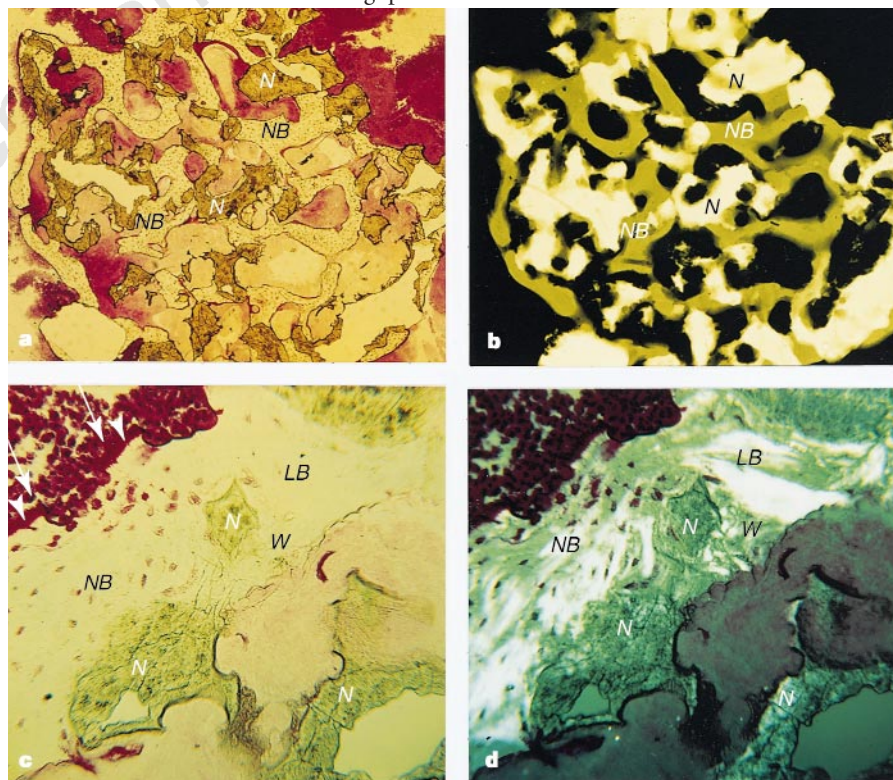


Figure 1 The *in vivo* effects of powdered nacre implant on bone regeneration in defective human jawbone, as revealed by Atlan *et al.*¹. The pictures are respectively histological (a, c – basic fuchsin staining), microradiographic (b) and polarized light (d) images of undecalcified sections of bone biopsies, six months after nacre (N) implantation. NB indicates newly formed bone. Comparison of the histological and the microradiographic images of the same section (a and b, respectively) show newly formed bone (low X-ray density) throughout the nacre (high X-ray density) implant. They are tightly welded together without intervening fibrous tissue, as shown in c. Polarized light imagery (d) indicates that the new bone is made of woven bone (W) and lamellar bone (LB) in close contact with the nacre implants. The bone surfaces are covered by osteoid borders (arrowheads in c) lined with osteoblasts, the bone-forming cells (arrows). a and b, $\times 26$ (from ref. 1); c and d, $\times 157$ (courtesy of E. Lopez).

Optical and electron microscopical studies revealed that the new tissue formed a continuous sheet spanning the space between the original fragments. Material near the original bone had all the characteristics of young, or 'woven', bone, whereas on the nacre side the characteristic brick-and-mortar structure of nacre was evident. *In situ* Raman spectroscopy showed that the newly formed nacre-like material was indeed aragonite, while apatite was found on the side of the bone fragment. So the new material was a mixed tissue, with nacre and bone coexisting. Control experiments using only osteoblasts, or osteoblasts and bone fragments placed 1 mm apart, did not induce any mineral growth or accretion. It was these experiments *in vitro* that encouraged Lopez and co-workers to carry out their daring implantation into patients¹.

Taken together, the results will provoke fresh controversy about basic issues in biomineralization. They point to an unexpectedly close kinship between very diverse animal skeletons, and raise the question of whether bone and nacre are homologous — that is, derived from incipient calcified skeletons in the common ancestor of molluscs

and vertebrates. The answer, in fact, is that they aren't. Reconstructions of early animal evolution based on DNA sequences of extant species, and on the fossil record, indicate that these skeletons evolved independently⁵. Indeed, although the first skeleton-forming animals lived contemporaneously with the famous soft-bodied Ediacaran biota in Late Neoproterozoic oceans⁶, some 570 million years ago, molluscan shells, vertebrate hard parts and most of the additional skeletal architectures seen in living animals evolved from 544 million years ago, as part of a broader Cambrian diversification of animal groups.

But although nacre and bone are not homologous as such, parts of the complex machinery that directs their formation may be. Logically, this would imply that, before the Cambrian diversification, these homologous parts served some function other than biomineralization and that they were co-opted for the latter purpose when the different metazoan stocks began producing skeletons. This picture is further supported by the observation that implants of coral skeletons — which are evolutionarily more distant from vertebrates than the molluscs — seem to be less effective than nacre in inducing bone formation in humans, although they, too, are biocompatible with human tissue⁷.

A first understanding of how such a calcifying machinery might have emerged from non-calcifying origins comes from the realization that organized calcification depends on suppression of uncontrolled crystal formation in supersaturated fluid compartments. For example, statherin proteins prevent spontaneous overcrusting of our teeth by supersaturated saliva⁸, and a glycoprotein in mice has been shown to prevent pathological deposits in a range of tissues⁹. Of particular relevance for early marine animals was the prevention of spontaneous encrustations on the outer body tissues in response to the highly oversaturated state of the Neoproterozoic ocean waters^{10,11}.

It seems, then, that unskeletonized Neoproterozoic animals strictly suppressed spontaneous crystallization from the pervading supersaturated solutions throughout their bodies. The subsequent emergence of calcification required the establishment of a sophisticated regulatory system that permitted crystallization to proceed locally and in a highly controlled and functional manner. The inhibitory molecules had to be inactivated or removed, and probably became incorporated into the mineral-inducing system, where they assumed a regulatory role in fine-tuning crystal growth^{12,13}. Concurrently, the constituent ions of the mineral had to be supplied in sufficient quantities to maintain adequate skeletonization, with macromolecules facilitating crystal nucleation¹⁴, growth and termination in appropriate locations¹⁵. Finally the organism had to

deploy signal transmitters able to switch this biomineralizing machinery on and off.

Remember that all of this machinery, the fine details of which are only superficially known, was put into place independently in diverse animal stocks and within a relatively short span of evolutionary time. This suggests that, in the face of ecological necessity¹⁶, different skeleton-forming systems were assembled at least in part from biochemical components inherited from common, uncalcified ancestors. The experiments of Lopez and her collaborators point to the signal transmitters as prime candidates for homologous components of bone tissue and nacre. A test of this hypothesis, now being pursued, lies in the identification, characterization and comparative analysis of these macromolecules. Particular attention is being paid to the identification of bone morphogenic proteins in nacre.

The idea behind this work did not appear out of the blue, for Mayan skulls discovered in 1931 have perfectly fitting teeth made of nacre and radiographs showed that their roots were perfectly integrated into the surrounding bone. Through the new approaches of Lopez and colleagues, it looks as if the old Mayan invention is about to give away its secret underpinnings. In so doing, these

approaches provide a potentially powerful experimental technique for exploring an association of functions that has been so conspicuously exploited for more than 500 million years. And it may result in new ways of treating bone deficiencies. □

Peter Westbroek and Frédéric Marin are in the Department of Chemistry, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands. e-mail: westbroek_p@rulgca.leidenuniv.nl

1. Atlan, G., Balmain, N., Berland, S., Vidal, B. & Lopez, E. *C. R. Acad. Sci. Paris/Life Sci.* **320**, 253–258 (1997).
2. Liao, H. et al. *J. Mater. Sci.* **8**, 823–827 (1997).
3. Lopez, E. et al. *Tissue Cell* **24**, 667–679 (1992).
4. Silve, C. et al. *Calcif. Tissue Int.* **51**, 363–369 (1992).
5. Bengtson, S. & Conway Morris, S. in *Origin and Early Evolution of the Metazoa* (eds Lipps, J. H. & Signor, P. W.) 447–481 (Plenum, New York, 1992).
6. Grant, S. W. F. *Am. J. Sci.* **290A**, 261–294 (1990).
7. Dupoirieux, L. et al. *Int. J. Oral. Maxillofac. Surg.* **23**, 395–398 (1994).
8. Raj, P. A. et al. *J. Biol. Chem.* **267**, 5968–5976 (1992).
9. Luo, G. et al. *Nature* **386**, 78–81 (1997).
10. Knoll, A. H. & Swett, K. *Am. J. Sci.* **290A**, 104–132 (1990).
11. Grotzinger, J. P. & Knoll, A. H. *Paleo* **10**, 578–596 (1995).
12. Marin, F., Smith, M., Isa, Y., Muzeyr, G. & Westbroek, P. *Proc. Natl Acad. Sci. USA* **93**, 1554–1559 (1996).
13. Wheeler, A. P., George, J. W. & Evans, C. A. *Science* **212**, 1397–1398 (1981).
14. Falini, G., Albeck, S., Weiner, S. & Addadi, L. *Science* **271**, 67–69 (1996).
15. Mann, S. *Struct. Bonding* **54**, 125–174 (1983).
16. Bengtson, S. (ed.) in *Early Life on Earth* 412–425 (Nobel Symp. No. 84, Columbia Univ. Press, New York, 1994).

Superconductors

Current limits to wire technology

David Christen

Soon after their discovery over a decade ago, high-temperature superconductors (HTS) promised applications in the power industry, where superconducting wires could carry large electric currents with virtually no energy loss. Unfortunately, the materials are extremely complex, and their properties present several serious obstacles to the development of wires. Today, many of these problems have been ingeniously solved, and 1-km lengths of BiSrCaCuO superconductors (BSCCO) are routinely produced by companies in the United States and Japan. These wires, which are actually multi-filament composite silver/superconductor ribbons, have already shown their practical uses in a number of prototype devices, including motors, transformers, fault-current limiters, and power transmission cables. But their performance is far below optimum, with less than 10% of the superconductor effectively carrying current. Understanding the sources of this problem is essential to the viability of the technology. On page 906 of this issue¹, Cai and colleagues describe experiments on single filaments extracted from a BSCCO tape, which shed new light on the barriers to current flow.

What are these barriers, and how can they be overcome? To appreciate the results of Cai

et al., it is helpful to look at the inherent obstacles posed by HTS materials. They are all multi-component compounds, composed of highly conducting CuO₂ sheets separated by ionic, nearly insulating layers. Superconducting currents flow freely within the CuO₂ sheets ('basal-plane' conduction), but along the crystalline *c* axis (perpendicular to the sheets) conduction only occurs by weak tunnelling of super-carriers through the ionic layers, and so in the *c* direction loss-free currents are very low.

The physical characteristics mimic the electronic, and the materials are brittle and layered, akin to many ceramics. For these reasons the early attempts to make wires by conventional 'toilet-bowl' ceramic technology — simply pressing and heat-fusing a powder of random polycrystals — yielded pitifully low current capacity. To make things worse, HTS compounds are oxides, and thus incompatible with the reactive metal copper that has traditionally been used as the support and stabilizing structure of low-temperature superconducting wires.

The BSCCO tape technology has gone a long way towards solving these problems. The tapes are made by loading a powder of BSCCO into a silver or silver-alloy tube, which is then sealed and drawn into a fine

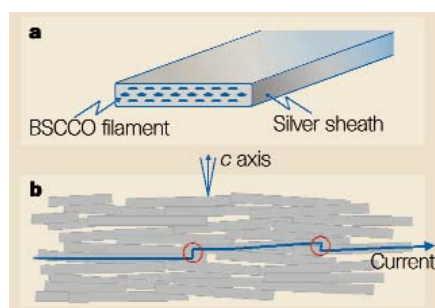


Figure 1 Sketches of (a) a standard industrial superconducting tape, with 'BSCCO' filaments embedded in a matrix of silver, and (b) the cross-section of a single BSCCO filament, tracing a flow of current through the assembly of crystalline grains.

wire. These wires are cut and re-stacked for a series of additional drawing, rolling and heat-treatment steps, leading to the final multi-filament oxide-powder-in-tube (OPIT) tapes (Fig. 1a). The silver matrix is chemically compatible with BSCCO, and endures the necessary high-temperature processing in oxygen. The fine BSCCO filaments, $\sim 100\text{ }\mu\text{m}$ wide by only a few μm high, tolerate an acceptable amount of bending, like the fine glass filaments in flexible fibre-optic cables. Most importantly, the processing aligns the crystalline BSCCO grains within a filament, so that the *c* axes are perpendicular to the tape plane (to within $10\text{--}15^\circ$). Within a filament, electric currents can thus flow from grain to grain predominantly within the CuO_2 sheets (Fig. 1b).

Why is the limiting current density still lower than observed in single crystals and epitaxial films? Cai *et al.*¹ have used magneto-optical imaging of magnetic flux penetration to show that many filaments are not active at all, owing to the intrusion of cracks from the tape edges. These gross defects are probably remnants of deformation and heat treatment, and might be cured by further refinements in processing. But the authors' measurements also revealed signs of some *c*-axis conduction, implying the existence of local barriers that block basal-plane conduction, forcing the weak, *c*-axis tunnel-currents to take over (Fig. 1b).

These barriers may be microstructural flaws, such as voids, cracks or second-phase defects. Or, more ominously, they may be intrinsic weak links, where the misalignment between adjacent grains is too large to accommodate basal-plane current transfer. Indeed, supercurrent transport across a grain boundary can be greatly suppressed² for misalignments greater than $\sim 10^\circ$. This is true even for the case where the *c* axes are perfectly parallel, and the adjacent crystals are merely misaligned in-plane (so-called [001] tilt-grain boundaries). This handicap may be intrinsic to HTS because of their low carrier density and strong sensitivity to structural disorder^{3–5}.

So, what are the prospects for improving BSCCO OPIT tapes? Refinements in thermo-mechanical processing — such as additional mechanical constraints, or higher-temperature rolling — should prevent filament cracks, and could also improve *c*-axis texture and filament microstructure. But processing trade-offs might persist, with competition between good interior filament properties and the elimination of gross defects near the tape edge. More seriously, at the moment there is no feature of OPIT tape processing that can lead both to *c*-axis and to in-plane grain alignment. Such biaxial alignment would give long conductors single-crystal-like properties, and work is under way to develop second-generation HTS wires by growing thin unbroken coatings of $\text{YBa}_2\text{Cu}_3\text{O}_7$ on flexible metal tapes⁶.

The obstacles may appear daunting, but the specifications of OPIT tapes have continued to improve steadily with time⁷. Although a ceiling awaits, elimination of the cracks, and improvements of a factor of two in the filaments would yield overall current densities of $\sim 50\text{ kA cm}^{-2}$ at liquid-nitrogen

temperatures. That is good enough for economically and environmentally appealing low-magnetic-field applications, such as underground transmission lines and power transformers. For this, the development would take about six years, but coated conductors may provide a leap in capability well before then. Indeed, commercial development of HTS is on schedule when compared with the histories of other high-technology products, such as integrated circuits and optical-fibre communications⁷. There is every reason to be optimistic that the current problems will be solved. □

David Christen is in the Solid State Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee 37831-6061, USA.
e-mail: dkc@ornl.gov

1. Cai, X. Y., Polyanski, A., Li, Q., Riley, G. N. Jr & Larbalestier, D. C. *Nature* **392**, 906–909 (1998).
2. Dimos, D., Chaudhari, P. & Mannhart, J. *Phys. Rev. B* **41**, 4038 (1990).
3. Gurevich, A. & Pashitskii, E. A. *Phys. Rev. B* (submitted).
4. Hilgenkamp, H. & Mannhart, J. *Phys. Rev. Lett.* (submitted).
5. Browning, N. D. *et al. Physica C* **294**, 183–193 (1998).
6. Hawsey, R. & Peterson, D. *Superconductor Industry* **9**, 23–29 (1996).
7. Yurek, G. *Superconductor Industry* **9**, 12–19 (1996).

Calcium signalling

Oscillation, activation, expression

Jacopo Meldolesi

I ncreased concentrations of free Ca^{2+} within the cytosol of living cells lead to the activation of arrays of both rapid and sustained events. These, and the events that cause them, can vary considerably under different conditions. As well as stable increases, most cells — both excitable and non-excitable — can develop oscillations in the concentration of cytosolic Ca^{2+} . These trains of Ca^{2+} spikes occur at much lower frequencies than those of membrane action potentials. But how do cells interpret these rises in the concentration of Ca^{2+} ?

Oscillations in the concentration of cytosolic Ca^{2+} have attracted considerable attention, and hypotheses of frequency coding were proposed a decade ago¹. But the experimental evidence has remained weak², and the physiological significance of the phenomenon is still questioned. On pages 933 and 936 of this issue, however, Dolmetsch *et al.*³ and Li *et al.*⁴ report highly original experiments in which oscillations were generated by new procedures, bypassing the complexities of cell heterogeneity and activation of surface receptors. The results clearly show that oscillations and their frequencies can be specific for gene activation, not only in terms of efficiency, but also of selectivity. Nothing of this kind has ever been demonstrated before, and these reports open a new field in Ca^{2+} signalling.

Li *et al.*⁴ report a great technological success — the development (by a 14-step

synthesis) of a modified version of inositol-1,4,5-trisphosphate, $\text{Ins}(1,4,5)\text{P}_3$, the second messenger that normally causes the release of intracellular Ca^{2+} stores. The newly developed molecule is not only membrane permeant but also caged, requiring photolysis (by ultraviolet light) for activation. It can therefore be loaded into intact, healthy cells, without inducing any effects until the $\text{Ins}(1,4,5)\text{P}_3$ is released. The unique advantage of this procedure is that the release of Ca^{2+} from intracellular stores (Fig. 1a, overleaf) can be modulated at will, in amplitude and in frequency, by varying the duration and frequency of the ultraviolet flashes.

The authors found that in RBL-2H3 basophilic leukaemia cells, trains of short (0.3–1.5-second) ultraviolet pulses, spaced 0.5–8 minutes apart, induced oscillations in the concentration of cytosolic Ca^{2+} . They then used this system to study, at the single-cell level, the activity of the Ca^{2+} -dependent transcription factor, nuclear factor of activated T cells (NF-AT). Activity of the gene was monitored using a procedure that the same group had developed⁵, based on transfecting RBL-2H3 cells with a reporter construct that drives the expression of β -lactamase. Li *et al.* found that oscillations in the concentration of intracellular Ca^{2+} were more effective than a single, prolonged increase, provided the period of each oscillation was roughly one minute — slower and faster frequencies were less efficient. In

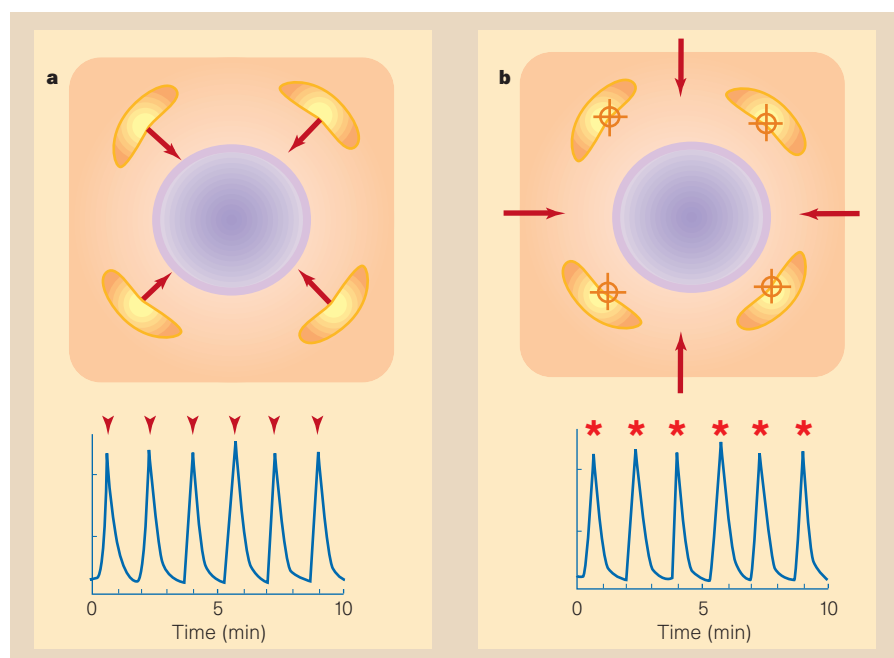


Figure 1 New procedures for generating homogeneous oscillations in the concentration of cytosolic Ca^{2+} . **a**, The method developed by Li *et al.*⁴. Cells are loaded with a caged version of inositol-1,4,5-trisphosphate, which is released by rhythmic flashes of ultraviolet light (arrowheads). Activation of the corresponding receptor in the endoplasmic reticulum leads to the release of Ca^{2+} to the cytosol (arrows) in an oscillating manner. **b**, The method developed by Dolmetsch *et al.*³. Irreversible blockade of SERCA, a Ca^{2+} pump in the endoplasmic reticulum, by thapsigargin ($\oplus$) prevents accumulation of Ca^{2+} within intracellular stores, so these become depleted. The resulting increase in the permeability of the plasmalemma to Ca^{2+} (due to activation of the store-dependent CRAC channels; arrows) is used to generate oscillations in the concentration of cytosolic Ca^{2+} . These oscillations are driven by rises in the concentration of Ca^{2+} in the incubation medium (asterisks).

the case of the faster frequencies, this may have been due to desensitization of the $\text{Ins}(1,4,5)\text{P}_3$ receptor⁴.

Dolmetsch *et al.*³ took a completely different approach — rather than stimulating $\text{Ins}(1,4,5)\text{P}_3$ -sensitive Ca^{2+} stores, they depleted them. Jurkat T cells were exposed to thapsigargin, which irreversibly blocks the specific Ca^{2+} pumps known as sarcoplasmic-endoplasmic reticulum Ca^{2+} -ATPases (SERCAs). This resulted in activation of the store-operated Ca^{2+} (CRAC) channels in the plasmalemma, leading to a stable increase in permeability of the cells to Ca^{2+} . Using this procedure, oscillations of controlled amplitude (generated by rapidly changing the concentration of Ca^{2+} in the incubation medium) were dynamically clamped. So, these oscillations were generated at the cell surface, without any involvement of intracellular receptors (Fig. 1b).

Dolmetsch *et al.* used this system to compare the effects, on NF-AT-induced gene expression, of oscillations and stable responses of similar average amplitude. They then looked, in parallel, at the activation of three Ca^{2+} -activated transcription factors — NF-AT, Oct/OAP and NF- κB . In each case, activation was monitored by the expression of reporter genes driven either directly or through the promoters of specific cytokines. Again, oscillations proved to be more effi-

cient than stable increases — but only when the average concentration of cytosolic Ca^{2+} remained below ~ 300 nM. Moreover (and even more excitingly), whereas the three transcription factors were activated in parallel with stable increases in the concentration of Ca^{2+} , such parallel activation was observed for the oscillations only if they occurred with a period of 400 seconds or less. With oscillations of a lower frequency, only NF- κB seemed to be stimulated. These low-frequency oscillations are thought to be triggered by intercellular signalling events, so they are important in cell physiology.

How can these results be explained? Activation of the three transcription factors studied by Dolmetsch *et al.* is mediated by a Ca^{2+} -dependent phosphatase called calcineurin. The transcription factors exist as complexes in the cytoplasm, and they are dephosphorylated by calcineurin. This induces dissociation of the complexes, followed by migration of the active subunits to the nucleus^{6,7}. The ensuing expression of specific genes is not controlled directly by Ca^{2+} , but it keeps going as long as the transcription factor remains in the nucleus. During oscillations, therefore, the work of transcription factors does not decline with the decreasing concentration of cytosolic Ca^{2+} — it can persist for longer. Moreover, because NF-AT returns to the cytoplasm



100 YEARS AGO

Striking is the difference in appearance between a Solpuga fasting and a Solpuga full fed. In the former the abdomen shrivels up, the segments shrinking one within another like the several pieces of a half-closed telescope; in the latter the expansion is carried to such an extent that the distended abdomen much resembles a short thick sausage, far surpassing in size and weight the rest of the body and limbs. This is brought about by the imbibition of water and of the fluid and semi-fluid tissues of their prey. In support of their water-drinking propensities, the following passage, written by the Sudan war correspondent to the *Standard* (October 19, 1897), may be cited: "One day in my tent [at Kerma] I heard a rustle like that of a silk dress. A big, ugly, yellow hairy beast, with nippers like a crab, was moving fast as a mouse over the moist ground near the zeer [porous water jar] in the corner of my tent. At last he settled down to suck the water from the sides of the jar." The writer of the passage just quoted had previously spoken of this animal as the "famous abu-shabat, the terror of the Sudan in the way of spiders, as large as your hand and ten times more venomous than a scorpion."

From *Nature* 28 April 1898.

50 YEARS AGO

Mainini has described (*Semana Medica*, 64, 337, March 1947) a pregnancy test in which, following injection of pregnancy urine into the male toad (*Bufo arenarum* Hensel), liberation of spermatozoa into the urinary bladder occurs within three hours; the same animal can be used again after five days. Drs. Octavio Rodrigues Lima and Oswaldo Gelli Pereira, of the Obstetrical Clinic, Medical School of Rio de Janeiro, University of Brazil, state in a communication submitted to the Editors that they have confirmed this work, using as test animal the male toad (*Bufo marinus*), as the species used by Mainini was not available to them. ... Sixty tests were carried out with the urine of amenorrhœic women. Positive results were found between half an hour and two hours, and were confirmed clinically in all cases. The test appears to be simple, economical and reliable.

From *Nature* 1 May 1948.

rapidly (~ one minute) after rephosphorylation by a specific kinase, whereas NF- κ B returns much more slowly (> 16 minutes after a single Ca^{2+} spike)^{4,8}, the persistent activity of NF- κ B during low-frequency oscillations can be explained.

The work of Dolmetsch *et al.*³ and Li *et al.*⁴ will make available to all Ca^{2+} laboratories direct procedures for generating oscillations in the cytosolic concentration of Ca^{2+} (Fig. 1), bypassing the drawbacks of heterogeneity due to cell and receptor variability. These studies will also enable us to begin to answer questions related to selectivity of Ca^{2+} signalling, and to look for subtle mechanisms of specificity in other Ca^{2+} responses. The general question is whether other signalling pathways are interpreted by cells — in terms of both the intensity and frequency of the

message — as shown here for calcineurin-dependent gene activation. □

Jacopo Meldolesi is in the Department of Pharmacology, University of Milan, and at DIBIT, Scientific Institute San Raffaele, Via Olgettina 58, 20132 Milan, Italy.

e-mail: meldolj@dibit.hsr.it

1. Berridge, M. J., Cobbold, P. H. & Cuthbertson, K. S. R. *Phil. Trans. R. Soc. Lond.* **320**, 325–343 (1988).
2. Tsien, R. W. & Tsien, R. Y. *Annu. Rev. Cell Biol.* **6**, 715–760 (1990).
3. Dolmetsch, R. E., Xu, K. & Lewis, R. S. *Nature* **392**, 933–936 (1998).
4. Li, W., Llopis, J., Whitney, M., Zlokarnik, G. & Tsien, R. Y. *Nature* **392**, 936–941 (1998).
5. Zlokarnik, G. *et al. Science* **279**, 84–88 (1998).
6. Crabtree, G. R. & Clipstone, N. A. *Annu. Rev. Biochem.* **63**, 1045–1083 (1994).
7. Timmerman, L. A., Clipstone, N. A., Ho, S. N., Northrop, J. P. & Crabtree, G. R. *Nature* **383**, 837–840 (1996).
8. Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C. & Healy, J. I. *Nature* **386**, 855–858 (1997).

Environmental science

Nitrogen oxides and tropical agriculture

A. F. Bouwman

Since 1992, and the appearance of the first scientific assessment of the Intergovernmental Panel on Climate Change¹, a great deal of evidence has accumulated showing that agricultural activities, especially application of fertilizers and animal manure, have resulted in increased emissions of nitrous oxide (N_2O) and nitric oxide (NO) to the atmosphere². The reasons why we should care about this are that N_2O is one of the so-called greenhouse gases, constituting 6% of the anthropogenic greenhouse effect², and also contributing to the depletion of stratospheric ozone; NO, too, is an important player in atmospheric chemistry, for it participates in regulating the oxidant balance in the troposphere.

These assessments are based on evidence from temperate areas. But what of the tropics? Writing in *Global Biogeochemical Cycles*³, Veldkamp, Keller and Nuñez discuss measurements^{3–6} of nitrogen oxide fluxes from soils in tropical agricultural systems and argue that fluxes in such systems may be much higher than those generally observed in temperate regions. They also point out that estimates of these fluxes depend less on fertilizer composition, as has sometimes been assumed, and more on the conditions in the soil concerned and the timing of treatment.

Release of nitrogen oxides occurs in soils during both biological nitrification and denitrification, and during chemical denitrification. Biological denitrification is the reduction of nitrate (NO_3^-) or nitrite (NO_2^-) to gaseous nitrogen oxides and molecular N_2 by essentially anaerobic bacteria. Both N_2O and NO can be produced and consumed by

denitrification. Chemical denitrification is the reduction of NO_2^- or NO_3^- by chemical reductants, while nitrification is the biological oxidation of ammonium (NH_4^+) to NO_2^- or NO_3^- under aerobic conditions.

After NO is emitted from the soil, it is often rapidly oxidized to nitrogen dioxide,

NO_2 , which is then readily absorbed onto leaf surfaces if a tree or shrub canopy is present, reducing the amount of NO and NO_2 escaping from the soil–plant system into the atmosphere. This reabsorption process is, however, less important in agricultural crops and grasslands than in dense forest canopies.

The regulation of the production and consumption of N_2O and NO by nitrification and denitrification in soil has been described by a ‘hole-in-the-pipe’ model (Fig. 1). A second model describes the effect of the soil–water status (Fig. 2). Veldkamp *et al.*³ found that the relationship between the water-filled pore space (WFPS) and N_2O fluxes observed in three studies^{3,4,6} in tropical systems was in close agreement with the theoretical model (Fig. 2), supporting the theory that in these systems NO stems primarily from nitrification and N_2O from denitrification.

Turning to look at differences in the crop systems, they found that, in tropical pastures with a fertile loamy soil developed from volcanic ash, the fraction of fertilizer nitrogen lost as N_2O (6.8%) and NO (1.3%) was much higher than the loss rates generally observed in temperate agricultural lands (about 1% and 0.5% for N_2O and NO, respectively^{7,8}). On a banana plantation with similar soil, the loss rates for N_2O and NO were 1.3–2.9% and 5.1–5.7%, respectively⁶. In a fertilized sugar-cane field in Hawaii⁴ and fertilized pastures with well-aerated soils in

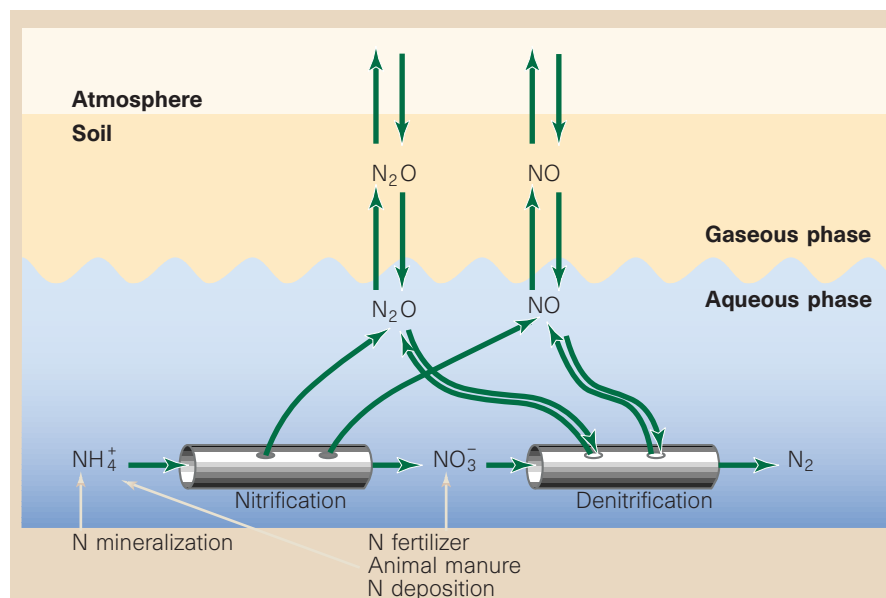


Figure 1 ‘Hole-in-the-pipe’ model of the regulation of trace-gas production and consumption by nitrification and denitrification¹². In this model, gas production and exchange with the atmosphere depend on: ● Factors controlling the amount of nitrogen flowing through the pipe (that is, those affecting denitrification and nitrification rates, mainly nitrogen availability and temperature). ● The size of the holes in the pipe through which nitrogen gases leak. Size is regulated by factors controlling the partitioning of the reacting nitrogen species to NO, N_2O or more reduced or oxidized forms, while the rate at which nitrogen moves through the pipes determines the importance of the leaks. ● The properties of the soil and length of the path between the production site and the open air. Before escaping from the soil to the atmosphere, the nitrogen gases diffuse through the soil pore system, where NO in particular may be taken up by plants or microorganisms.

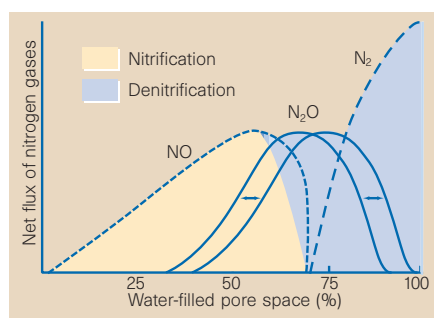


Figure 2 Model of the relationship between water-filled pore space (WFPS) of soils and the relative fluxes of nitrogen gases from nitrification and denitrification¹². According to this model, the highest NO fluxes are expected at 30–60% WFPS, when nitrification is most active, and the highest N₂O fluxes at 50–80% to 60–90% WFPS (depending on the soil properties), when denitrification dominates. Above 80% WFPS, oxygen becomes so limited that N₂O is consumed by denitrification and N₂ becomes the main end product.

Puerto Rico⁵, however, the percentage of fertilizer lost as N₂O came close to the loss rates generally observed in temperate regions.

These differences in fluxes stem from a combination of management and environmental conditions. In the sugar-cane field, application rates of fertilizer were adjusted to plant demand, lowering the nitrogen available to feed the process pipes (Fig. 1). On the banana plantation, by contrast, excess nitrogen was higher, thereby allowing a large amount of it to flow through the pipes and making oxygen the dominant controlling factor. As the pastures studied by Veldkamp *et al.*³ generally showed a higher percentage of WFPS, conditions for N₂O formation (denitrification) were more favourable and those for NO production (nitrification) less favourable than on the banana plantations and in the sugar-cane fields (Fig. 2).

The authors also found³ that the effect of fertilizer composition (NH₄⁺, urea and NO₃⁻, each combined with triple superphosphate, and also urea only) on N₂O and NO emissions was not significant. The soil water content and the timing of fertilization had far greater influence than fertilizer composition.

The broader environmental context of these results is that high rates of application of nitrogen fertilizer are common in the developing, mainly tropical countries, and are likely to increase. A global inventory of synthetic fertilizer use by crop indicates an average nitrogen application rate in developing countries of only 57 kg N ha⁻¹ in 1990, compared with 86 kg N ha⁻¹ in the industrialized countries in the same year⁹. But in developing countries the fertilizer is applied to only a fraction of the agricultural land¹⁰, implying that the application rate is much higher than it would seem from the average

rate. Except for rice cultivation, most agricultural production in the tropics generally takes place on well-drained soils. So the conditions of high rates of nitrogen fertilizer use, and the aerobic state and largely unrestricted diffusion in the soils concerned (Fig. 1), favour large gas fluxes — thus supporting the view of Veldkamp *et al.*³ that tropical agricultural land is potentially a large emitter of N₂O and NO.

This conclusion has implications for the global budget of gaseous nitrogen oxides. Based on measurements from temperate agriculture only, it is estimated that agricultural sources contribute more than 80% of the current annual increase in atmospheric N₂O (ref. 2). The importance of tropical agriculture in the N₂O budget, and in tropospheric chemistry through NO emissions, is likely to increase in the future, given that use of nitrogen fertilizers increased by almost 20% in the first half of this decade, from 42 million tonnes of nitrogen in 1990 to 50 million tonnes in 1995 (ref. 9). Projections also indicate an increase of up to about 80 million tonnes of nitrogen in the year 2010 and to 106 million tonnes in 2025 (ref. 11).

Finally, Veldkamp and colleagues make a proposal, stemming from their finding of the sensitivity of gas fluxes to soil moisture and timing of fertilizer use, that should alter

thinking in field experiments in this area of research. They suggest that site selection for flux measurements should be based on differences in soil moisture regimes and fertilizer management, enabling investigation of correlations between fluxes of nitrogen oxides and these key factors. □

A. F. Bouwman is in the National Institute of Public Health and the Environment, PO Box 1, 3720 BA Bilthoven, The Netherlands.

e-mail: Lex.Bouwman@rivm.nl

1. Houghton, J. T., Callander, B. A. & Varney, S. K. (eds) *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment* (Cambridge Univ. Press, 1992).
2. Houghton, J. T. *et al.* *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios* (Cambridge Univ. Press, 1995).
3. Veldkamp, E., Keller, M. & Nuñez, M. *Glob. Biogeochem. Cycles* **12**, 71–79 (1998).
4. Matson, P. M. *et al.* *J. Geophys. Res.* **101**, 18533–18545 (1996).
5. Mosier, A. R. & Delgado, J. A. *Chemosphere* **35**, 2059–2082 (1997).
6. Veldkamp, E. & Keller, M. *J. Geophys. Res.* **102**, 15889–15898 (1997).
7. Bouwman, A. F. *Nutrient Cycling Agroecosys.* **46**, 53–70 (1996).
8. Veldkamp, E. & Keller, M. *Nutrient Cycling Agroecosys.* **48**, 69–77 (1997).
9. <http://apps.fao.org/default.htm>
10. *Fertilizer Use by Crops 2, Rep. ESS/MISC/1994/4* (Food and Agriculture Organization of the United Nations, Rome, 1994).
11. *Land and Water Bull. No. 6* (Food and Agriculture Organization of the United Nations, Rome, 1997).
12. Davidson, E. A. in *Microbial Production and Consumption of Greenhouse Gases: Methane, Nitrogen Oxides and Halomethanes* (eds Rogers, J. E. & Whitman, W. B.) 219–235 (Am. Soc. Microbiol., Washington DC, 1991).

Ornithology

The purple patch

Tim Guilford and Paul H. Harvey

It is hard to sex a blue tit — even experienced ornithologists regard males and females as essentially the same in coloration. Papers by Andersson *et al.*¹ and Hunt *et al.*² in *Proceedings of the Royal Society* now show that blue tits are, in fact, sexually dichromatic. The reason that this was not known many years ago is that the newly recognized colour differences are in the ultraviolet part of the colour spectrum,

which birds can detect³ but we and other mammals cannot. This is not the first demonstration that birds can show mate selection behaviour based on variation in the ultraviolet^{4,5}. But it is the first instance in which it has been shown that a bird species that is colourful to humans is clearly sexually dichromatic only in the spectrum that is invisible to humans.

The two groups worked on the same



Figure 1 Male or female? We can't tell, because blue-tit sexes are virtually indistinguishable to the human eye. As Andersson *et al.*¹ and Hunt *et al.*² show, however, the birds themselves can detect a striking ultraviolet dimorphism, a purple patch as it were, on the crest at the top of the head.

A. T. D. BENNETT & S. HUNT

species of tit, *Parus caeruleus*, but in Sweden and England respectively. Andersson *et al.*¹ used reflectance spectrophotometry to measure the colour of the tits' crown feathers, which are slightly erected during display. The sexes differ significantly in spectral shape around the ultraviolet, with males showing purer colours and a shorter-wavelength peak (430 nm as opposed to 443 nm in females). Because colour perception is influenced by the contrast with background colours, Andersson *et al.* computed the spectral contrast of the crown with natural backgrounds. Against a background of dead leaves, typical of early spring when mate choice takes place, spectral contrast was both maximized and most sexually dimorphic at 370 nm, right at the peak sensitivity of the ultraviolet receptor in the passerines⁶ (of which blue tits are a member).

Hunt *et al.*² measured colour for ten regions of the body, and found that spectral shape was dimorphic in half of them. Again, though, the crest proved the most interesting region, with a major ultraviolet dimorphism and shorter-wavelength peak reflectance in males (349 nm against 355 nm in females). The lack of correspondence between the actual peaks in the two studies may be a sampling artefact, but, more intriguingly, it could represent a genuine difference between the colours of Swedish and English blue-tit populations.

Both groups also demonstrated that tits not only look different from one another, they also care about the difference. Andersson *et al.* found that wild Swedish pairs were assortatively mated for crown colour — that is, paired individuals were more similar to each other in crown colour than expected by chance. Hunt *et al.* showed in the laboratory that English females preferred to associate with those males that have the brightest crowns.

Blue tits are signalling to one another in the ultraviolet, but we cannot see that signal without the aid of sophisticated equipment. Why, then, are blue tits so strikingly coloured (Fig. 1), yet largely monomorphic in our visible spectrum? Is it a coincidence that these birds have evolved to reveal their dimorphisms only to other birds, which seem in general to possess ultraviolet vision³, and not to mammals? Those interested in the evolution of bird coloration might interpret this new finding as a possible vindication of Baker and Parker's⁷ claim that bright coloration in birds has evolved as a signal to predators, in this case mammalian predators. Their argument was that bird species that are unprofitable prey, perhaps because they are good at escaping, should advertise the fact honestly by having striking coloration. Where both sexes are equally unprofitable as prey, monomorphic bright coloration would be favoured. The monomorphic striking coloration that is

apparent to mammals would, according to this view, be the optimum coloration for avoiding mammalian predators. The dimorphism is just for the birds.

However, the main natural predators of adult blue tits are almost certainly other birds, such as sparrowhawks (*Accipiter nisus*) — which, if ultraviolet vision is indeed a general property of birds³, would be well able to detect the sexual dimorphism. What is more, small variations in colour patterns that are visible to mammals have been shown to be used as a social signal between individuals in the closely related great tit (*Parus major*)⁸, which has similarly striking coloration and intricate plumage patterning. It seems, then, that these two studies^{1,2} pose a new problem. If the important receivers of blue-tit signals are birds — whether predators, rivals or mates — then why are the monomorphic signals restricted to one part of the spectrum

(which happens to be visible to mammals), and the dimorphic signals restricted to another part of the spectrum (which happens to be invisible to mammals)? □

Tim Guilford and Paul H. Harvey are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.
e-mail: tim.guilford@zoology.oxford.ac.uk
paul.harvey@zoology.oxford.ac.uk

1. Andersson, S., Örnberg, J. & Andersson, M. *Proc. R. Soc. Lond. B* **265**, 445–450 (1998).
2. Hunt, S., Bennett, A. T. D., Cuthill, I. C. & Griffiths, R. *Proc. R. Soc. Lond. B* **265**, 451–455 (1998).
3. Bennett, A. T. D. & Cuthill, I. C. *Vis. Res.* **34**, 1471–1478 (1994).
4. Bennett, A. T. D., Cuthill, I. C., Partridge, J. C. & Maier, E. J. *Nature* **380**, 433–435 (1996).
5. Hunt, S., Cuthill, I. C., Swaddle, J. P. & Bennett, A. T. D. *Anim. Behav.* **54**, 1383–1392 (1997).
6. Bowmaker, J. K., Heath, L. A., Wilkie, S. E. & Hunt, D. M. *Vis. Res.* **37**, 2183–2194 (1997).
7. Baker, R. R. & Parker, G. A. *Phil. Trans. R. Soc. Lond. B* **287**, 63–130 (1979).
8. Wilson, J. D. *Anim. Behav.* **43**, 999–1009 (1992).

Neuropharmacology

Premenstrual steroids?

Karen T. Britton and George F. Koob

Premenstrual syndrome (PMS) is a disorder that is characterized by depression, anxiety and mood swings. It regularly occurs during the last week of the luteal phase of the menstrual cycle, two weeks before menses^{1,2}, and the symptoms have a considerable negative effect on the social and occupational lives of severely afflicted women. The menstrual cycle is associated with large fluctuations in a variety of steroids³ that are thought to be involved in PMS⁴, and, on page 926 of this issue, Smith *et al.*⁵ provide an insight into the potential link between members of one such group of steroids — the neuroactive steroids — and an animal model of PMS.

The concept of 'neuroactive steroids' was introduced by Baulieu⁶, and the term refers to steroids that are synthesized *de novo* from cholesterol and act on the brain³. These compounds include dehydroepiandrosterone (DHEA) and pregnenolone, as well as their esters (DHEA sulphate and pregnenolone sulphate, respectively), progesterone and the 5 α -reduced progesterone metabolite, allopregnanolone (3 α -OH-5 α -pregnan-20-one; Fig. 1, overleaf). Neuroactive steroids are distinguished from hormonal steroids by their neuropharmacological potency, and structure–activity requirements at membrane receptors — hormonal steroids act at cytosolic receptors.

Some neuroactive steroids, such as allopregnanolone, can produce behavioural effects that are similar to those caused by benzodiazepine drugs, including anti-convulsant⁷ and anxiolytic-like⁸ actions. They alter neuronal excitability by binding to allosteric sites on neurotransmitter-gated

ion channels, such as the γ -aminobutyric acid (GABA)-mediated chloride ion channel. This binding leads to either inhibition or potentiation of the inhibitory effects of GABA on this channel. Inhibition of GABA-receptor function produces effects ranging from anxiety and excitability to seizure activity; potentiation, by contrast, produces anxiolytic actions, sedation and sleep.

The effects of some neuroactive steroids, such as allopregnanolone and alloTHDOC (3,21-dihydroxy-5-pregnan-20-one), fluctuate with ovarian and adrenal function^{3,9}. For example, allopregnanolone and pregnenolone are both anxiolytic metabolites of progesterone. Circulating concentrations of these hormones are, therefore, highly correlated with levels of progesterone during the human menstrual cycle^{4,10,11}. Premenstrual symptoms such as anxiety and susceptibility to seizures are associated with precipitous decreases in circulating levels of progesterone, and similar effects have also been observed after withdrawal from benzodiazepines and alcohol. Consistent with changes in hormonal function throughout the menstrual cycle, women with PMS become insensitive to the actions of benzodiazepines¹² and allopregnanolone¹³ during the luteal phase, and they are also more susceptible to seizures¹⁴. Moreover, if female rats are given chronic levels of progesterone and the hormone is then abruptly withdrawn, they show similar symptoms — insensitivity to the actions of benzodiazepine drugs and an increased susceptibility to seizures.

In a series of elegant experiments, Smith and colleagues⁵ now show why these effects occur. They find that the mechanism re-

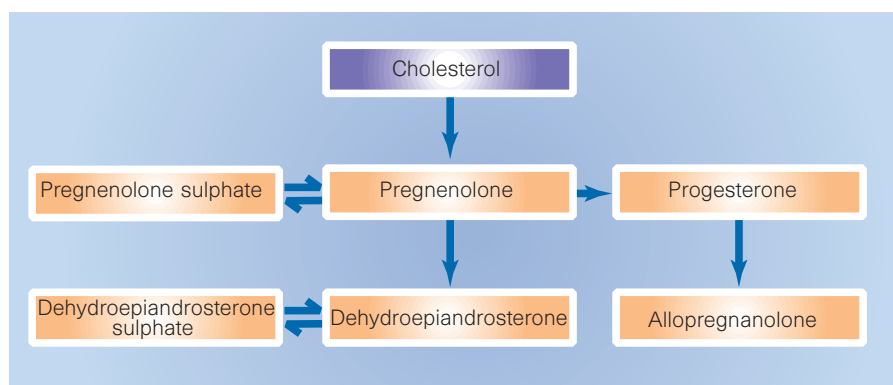


Figure 1 Relationships between some of the known neuroactive steroids. This group of steroids is synthesized from cholesterol. Smith *et al.*⁵ have found that, in female rats, decreased production of allopregnanolone — due to decreasing levels of progesterone during menstruation — leads to increased production of the $\alpha 4$ subunit of the γ -aminobutyric acid GABA_A receptor. This changes the sensitivity of the GABA_A receptor to endogenous ligands, resulting in symptoms associated with premenstrual syndrome, such as increased susceptibility to seizures and insensitivity to benzodiazepine drugs.

sponsible seems to be the abrupt decline in progesterone and allopregnanolone, and that this leads to changes in production of a specific subunit of the GABA_A receptor. The authors used indomethacin to block the formation of allopregnanolone in female rats that were exposed to progesterone. When they then withdrew the progesterone, they found that the rats did not show the insensitivity to benzodiazepine that is usually seen. They also observed an increase in levels of the $\alpha 4$ subunit of the GABA_A receptor, both *in vitro* and *in vivo*, following progesterone withdrawal. Moreover, when they treated rats with antisense oligonucleotides against the $\alpha 4$ subunit (to prevent transcription of the protein), this also reversed the benzodiazepine insensitivity and increased susceptibility to seizures that are usually seen during progesterone withdrawal.

These results indicate that fluctuations in endogenous levels of progesterone — via its metabolite allopregnanolone — may result in changes in the sensitivity of GABA_A receptors to endogenous ligands, resulting in an increased susceptibility to seizures and insensitivity to benzodiazepines. Because such fluctuations in neuroactive steroids occur in the menstrual cycle and in pregnancy³, they may be responsible for some of the symptoms observed. But Smith and colleagues' study represents only a part of an explosion of interest in modulation of brain function by neuroactive steroids. Many steroids have been identified that have the potential to act like neuroactive steroids, some related to corticosteroids and others to progesterone. Such steroids may be found in the brain, independent of circulating hormones⁶. Moreover, steroids such as pregnenolone sulphate and DHEA sulphate have been shown to modulate glutamate receptors^{15,16}, and they may play a role in memory.

This neuroactive-steroid connection may

prove to be involved in sedative-hypnotic actions⁸, ageing¹⁷, stress¹⁸ and alcohol abuse¹⁹. Perhaps more importantly, the steroids derived from progesterone may explain a variety of symptoms during pregnancy and the menstrual cycle, and, perhaps, some of the differences between men and women in the incidence of anxiety and mood disorders¹. □

Karen T. Britton is in the Department of Psychiatry, Veterans Administration Medical Center and University of California, San Diego, California 92161, USA.

e-mail: kbritton@ucsd.edu

George F. Koob is in the Department of Neuropharmacology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

e-mail: gkoob@scripps.edu

1. American Psychiatric Association *Diagnostic and Statistical Manual of Mental Disorders* 4th edn (Am. Psychiat. Press, Washington DC, 1994).
2. Bloch, M., Schmidt, P. J. & Rubinow, D. R. *Am. J. Psychiat.* **154**, 1741–1746 (1997).
3. Paul, S. M. & Purdy, R. H. *FASEB J.* **6**, 2311–2322 (1992).
4. Wang, M., Seippel, L., Purdy, R. H. & Blackstrom, T. J. *Clin. Endocrinol. Metab.* **81**, 1076–1082 (1996).
5. Smith, S. S. *et al. Nature* **392**, 926–930 (1998).
6. Baulieu, E. E. in *Steroid Hormone Regulation of the Brain* (eds Fuxe, K., Gustafsson, J. A. & Wetterberg, L.) 3014 (Pergamon, New York, 1981).
7. Gee, K. W., McCauley, L. D. & Lan, N. C. *Crit. Rev. Neurobiol.* **9**, 207–227 (1995).
8. Brot, M., Akwa, Y., Purdy, R. H., Koob, G. F. & Britton, K. T. *Eur. J. Pharmacol.* **325**, 1–7 (1997).
9. Majewska, M. D., Harrison, N. L., Schwartz, R. D., Barker, J. L. & Paul, S. M. *Science* **232**, 1004–1007 (1986).
10. Schmidt, P. J., Purdy, R. H., Moore, P. H., Paul, S. M. & Rubinow, D. R. *J. Clin. Endocrinol. Metab.* **79**, 1256–1260 (1994).
11. Rapkin, A. J. *et al. Obs. Gynaecol.* **90**, 709–714 (1997).
12. Sundstrom, I., Nyberg, S. & Blackstrom, T. *Neuropsychopharmacology* **17**, 370–381 (1997).
13. Sundstrom, I. *et al. Neuroendocrinology* (in the press).
14. Zimmerman, A. W. *Neurol. Clin.* **4**, 853–861 (1986).
15. Flood, J. F., Morley, J. E. & Roberts, E. *Proc. Natl Acad. Sci. USA* **92**, 10806–10810 (1997).
16. Park-Chung, M. *et al. Mol. Pharmacol.* **52**, 1113–1123 (1997).
17. Vallee, M. *et al. Proc. Natl Acad. Sci. USA* **94**, 14856–14870 (1997).
18. Barbaccia, M. L. *et al. Br. J. Pharmacol.* **120**, 1582–1588 (1997).
19. Devaud, L. L., Purdy, R. H. & Morrow, A. L. *Clin. Exp. Res.* **19**, 350–355 (1995).

Daedalus

The wide open society

Last week Daedalus unveiled his 'Lying Eye' video-analysis computer system, for telling if a speaker is lying. It decodes the body language of many small facial, bodily and verbal clues to spot the subtle signature of dishonesty, and gauges its magnitude. It will transform video conferences and law courts — and much else besides. For deceit and concealment are universal social skills. The human subconscious itself may have evolved as a safe place in which to hide the truth, freeing the conscious mind to deploy its own lies, and detect the lies of others. A machine that spots deception will transform society. Hysteria, that fashionable Victorian syndrome, vanished in this century through being 'rumbled'. Daedalus hopes that Lying Eye will similarly expose many current emotional fashions.

Its first target will be the indignation industry. Claims to be shocked or offended, accusations of insensitivity, harassment or abuse, gusts of conspicuous compassion towards socially approved underdogs, all will wither under the cool scrutiny of the Lying Eye video scanner. With any luck, a whole portfolio of self-righteous posturing and virtuous outrage will be thoroughly shown up. Even the keenest players will have to abandon the game.

But Lying Eye itself is only the prototype of a broader emotion-detector. It makes its deductions from a vast wealth of non-verbal data — a full spatial Fourier analysis of all face and body movements, correlated with audio output. Many other emotions must also be coded in these data: hostility, superiority, nervousness, anger, drunkenness or druggedness, sexual invitation or intent, criminality, and so on. An improved video analyser could detect them all. Daedalus is developing a wider and more detailed program, provisionally code-named 'Insight', for the job.

'Insight' will initially be aimed at psychiatry and counselling. Clients will be happy to have their basic problems and attitudes identified at once, and their progress under therapy accurately monitored. But Insight should also soon escape into wider society. It should expose trickery and game-playing of all kinds, and enforce far more authentic and healthy styles of social interaction. Ultimately, Insight will be fitted to every security camera. Social harmony will then reach its final peak. Big Brother will know the location, feelings and intentions of each of his loyal subjects all the time.

David Jones

Molecular clocks: mastering time by gene regulation

Paolo Sassone-Corsi

Self-sustaining clocks that regulate daily and seasonal rhythms are found in many biological systems, from fungi to humans. The structure and function of the molecular gears that control these clocks through the finely tuned regulation of gene expression are now being unravelled.

Imagine that you're locked in a room with no clocks and no windows. For most people, the cyclic patterns of sleeping, waking, activity and hunger will soon adopt a period of about 23 hours — evidently, the body 'knows' the length of the day, even in the absence of visual cues. Environmental signals, such as the day–night cycle, simply modify this biological rhythm to a slightly longer period of 24 hours.

In mammals, this internal clock controls rhythmic physiology and behaviour such as hormone secretion, breeding cycles and locomotor activity (for example, running, scratching and rearing in mice). In plants, a similar clock regulates daily oscillations in photosynthesis; and, in fruitflies, a molecular clock regulates emergence from the pupal case and locomotor activity¹.

The clock is composed of molecular gears, and recent studies point to a role for transcription (the process by which the DNA template is read to produce a messenger RNA transcript) in measuring time². Proteins defined as 'transcription factors' bind specifically to DNA and to other proteins, changing the pattern of gene expression and thus leading to responses that modulate the function of the clock. But where is the clock? When does it start ticking? And how does it work?

What is the clock?

Until the 1930s, biological rhythms were studied only by botanists. In 1729, the French astronomer d'Ortous de Mairan reported that the leaves of a heliotrope plant could open and close following a day and night rhythm. Then, 100 years later, Augustin de Candolle showed that these leaves could also open in constant darkness, indicating that the cycle was independent of the light–dark rhythm.

Research has since broadened to include almost every organism with a nucleus (eukaryotes) as well as some bacteria, and we now know that the clock consists of three conceptual components (Fig. 1). First, there is an input pathway that links the internal cycle to

external light–dark patterns. Second is the autonomous pacemaker, which generates the daily (circadian) oscillation. Finally, there is an output pathway for the expression of the physiologically measurable rhythms. For many years it was thought that the control of such complex and subtle regulations required tissue organization and high-level intercellular communications. In fact, every cell in the pacemaker constitutes an autonomous clock, with its own oscillatory properties^{1,2}.

Where is the clock?

In many organisms, the clock is localized in specific areas of the central nervous system. The pineal gland of birds, reptiles and fish, for example, contains photosensitive cells that show pacemaker properties and direct the circadian production of hormones. In mammals the situation is more complicated. The centre of the clock is an area in the anterior hypothalamus of the brain, called the suprachiasmatic nucleus (SCN). The SCN contains about 10,000 densely packed neurons, which are among the smallest cells in the brain. These neurons secrete a number of neuropeptides (such as the hormone vasopressin) with a circadian rhythmicity. They also send rhythmic signals to the pineal gland, which translates these signals into the circadian production of another hormone, melatonin^{3,4}.

Among the neural inputs that feed into the SCN are visual cues from the retina, which seem to be both necessary and sufficient in mammals for coupling circadian

rhythms to the light–dark cycle. Evidence that the SCN is the mammalian pacemaker includes the fact that electrical or pharmacological stimulation of this region causes circadian rhythms in the rat to shift phase. It has also been shown in hamsters that surgically removing over 75% of the SCN eliminates rhythmicity in locomotor activity, food intake, temperature control and the secretion of hormones such as melatonin and prolactin. After this treatment, rhythmic function is never recovered. But it can be re-established by giving the hamsters SCN neural grafts, and the restored rhythms show properties that are characteristic of the circadian pacemaker of the donor, rather than that of the host⁵.

Is the SCN a master clock controlling all systems that oscillate with a circadian periodicity in mammals? Probably not. A genetically programmed circadian oscillator has been found in the retina of the golden hamster. Although this activity is self-sustaining for several days in constant darkness, it is also sensitive to light — similar to the pacemaking activity of the chick pineal gland. This indicates that the hamster retinal system is independent of the SCN⁶. In the fruitfly *Drosophila melanogaster*, the malpighian tubules — which act like a kidney — express specific genes according to a cyclical daily rhythm. But experiments using transgenic flies that express either luciferase (which glows) or green fluorescent protein have shown that autonomous circadian oscillators are present throughout the fly's body⁷. This suggests that individual cells can support their own independent clocks. It is not known whether the activity of these diversely spread clock cells is coordinated in some manner or, if so, how. Perhaps there are various clocks for different physiological situations, residing in distinct anatomical structures, and such parallel clocks might overlap to regulate the output.

Clock genes

Clock molecules have been isolated in various organisms — from the fungus *Neurospora crassa* to *Drosophila* to the mouse — and always with the help of genetics (Table 1, overleaf). The first evidence for a molecule involved in controlling the clock came from the fortuitous isolation of a spontaneous

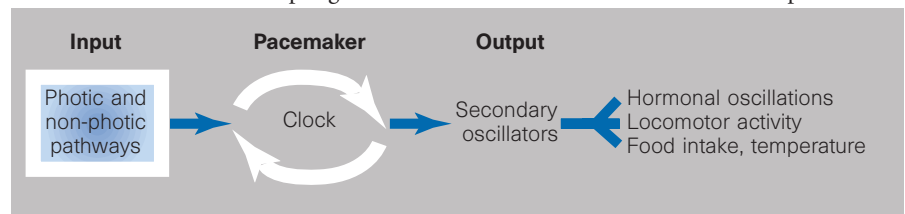


Figure 1 The components of a molecular clock. The clock receives input signals that connect it to geophysical cues such as circadian (day–night) rhythms. The clock itself (also known as the pacemaker) is an autonomous oscillator, made up of a molecular mechanism that is found within individual cells — interactions between cells are not required. The clock directs peripheral circadian phenomena (outputs), possibly through secondary oscillators.

mutation, *tau*, in the hamster. Animals with one copy of the mutated gene show a 22-hour period in locomotor activity, and animals with two mutated copies have a 20-hour periodicity⁸. Unfortunately, further molecular analysis of the *tau* locus has not been easy, owing to the lack of well-developed genetic maps in hamsters.

The mouse, however, is genetically well characterized, and a number of genes have been identified that are rapidly induced by light in its SCN⁹. All of these genes encode transcription factors that contain structural motifs such as the leucine zipper, which mediates protein–protein interactions, and the zinc finger, which mediates protein–DNA interactions. Expression of some of these genes oscillates during the light–dark cycle², but it is difficult to work out where these genes fit into the clock mechanism — because a clock can both measure and show time, those genes that are directly implicated in the clockwork need to be distinguished from those that mediate output signals.

In *Neurospora*, the endogenous clock is responsible for the precise circadian synchronization of cycles of asexual sporulation (conidiation), and the *frequency* (*frq*) gene seems to be a central component of the oscil-

lator¹⁰. Transcription to produce *frq* messenger RNA is central to the clock, with levels of the *frq* transcript peaking in the morning. If an inducible *frq* gene is expressed at the wrong time, the conidiation rhythm is abolished; conversely, suppression of the inducible gene resets the conidiation cycle.

The *frq* messenger RNA transcript can be translated into protein from either of two initiation sequences (codons), resulting in the production of two Frq polypeptides¹¹. Changes in temperature regulate the choice of codon used, setting the physiological temperature limits for rhythmicity. Immediately after synthesis, phosphate groups are added to both proteins (phosphorylation), yielding several forms of Frq¹². It is not clear how Frq contributes to the generation of circadian rhythmicity, but it is probably at the level of transcription — indeed, Frq needs to be in the nucleus (where transcription occurs) to be active¹⁰. Moreover, although Frq does not belong to any well-defined class of transcription factor, it has several features that are typical of such factors: these include a putative helix–turn–helix domain through which it can bind DNA; a signal that allows it to be targeted to the nucleus (a nuclear-localization signal); and highly charged regions that

could be involved in transcriptional activation (Table 1).

Association of clock molecules

In *Drosophila*, at least two genes are involved in clock function, and the dynamic association between their products ensures correct circadian rhythmicity. One of these, *period* (*per*), was the first clock gene ever to be isolated¹³. Nonsense mutations in *per* (which lead to an abnormally shortened Per polypeptide) result in the loss of rhythmic locomotor activity in *Drosophila*. Long-lasting phase shifts in locomotor activity are also observed when pulses of heat are given to transgenic flies bearing a heat-inducible copy of *per*. Missense mutations, which cause single amino-acid substitutions in Per, change the length of the circadian cycle^{2,14}.

Cyclical expression of *per* is regulated by transcription. The promoter region of the *per* gene, which is required for initiation of transcription, is sufficient to confer oscillating transcriptional regulation on another gene¹⁴. But production of a stable messenger RNA transcript would lead to the accumulation of this *per* messenger RNA and, hence, only minor oscillations. The trick that the *per* gene uses to ensure a strict rhythm is to have a relatively short messenger RNA half-life¹⁵. This is likely to be a very efficient regulatory mechanism, probably involving the rhythmic activation of factors involved in RNA processing. A mechanism of this kind exists in the dinoflagellate *Gonyaulax polyedra*, for example — circadian expression of a gene correlates with the cyclic binding of a protein to the 3' end of the messenger RNA, regulating the stability of the transcript¹⁶.

The second fly clock gene is *timeless* (*tim*). Mutations in *tim* have dramatic consequences on *per* (Fig. 2): the oscillatory expression is lost; phosphorylation of the Per protein is disrupted; and time-dependent nuclear transport is abolished. Conversely, cyclic expression of *tim* is altered in *per* mutants¹⁷. The Tim protein seems to regulate Per by interacting with it to form a Per–Tim dimer. Per contains a region called the PAS domain (Table 1), which is necessary for protein–protein interactions¹⁸, and is now thought to be the 'signature' of this class of clock molecule. PAS domains were named after the three proteins in which this structural motif was first identified: *Drosophila* Per, the mammalian Arnt, and Sim, which is the product of the fly *single-minded* gene. These domains have since been found in several transcription factors, where they are thought to confer target-gene specificity, and they are often coupled to a DNA-binding domain (a basic helix–loop–helix or bHLH domain). *Drosophila* Per does not contain a bHLH domain, however, suggesting that it may regulate transcription without binding DNA.

Surprisingly, Tim does not seem to con-

Molecular clocks in development

During the formation of an embryo, the events that lead to determination of the various cell types seem to be controlled by developmental timers which operate within individual cells. A remarkable example of this timekeeping is provided by the transition from rapid and symmetrical cell divisions to slow and asymmetrical divisions in embryos of the toad *Xenopus laevis*. This transition always occurs at the twelfth cleavage after fertilization, when the embryo consists of just over 2,000 cells known as blastomeres. The timing of the transition is controlled by a clock that is intrinsic to each blastomere — timing does not depend on cell–cell interactions³³. Another example comes from precursor cells that differentiate to form

oligodendrocytes, a type of non-neuronal cell found in the vertebrate nervous system. When placed individually in single wells, these precursors divide a number of times before differentiating. Two daughter cells from the same precursor undergo a synchronized number of divisions, and differentiate at the same time³⁴.

Developmental clocks such as these are probably distinct from those that govern circadian rhythms, although the molecular mechanisms that control proliferation and differentiation programmes are obviously intermingled with cellular oscillatory functions. For example, expression of the avian equivalent of the *Drosophila* gene *hairy* in the chick is controlled by a molecular clock that is

linked to the formation of somites. *Hairy* is expressed in cyclic pulses with a periodicity of 90 minutes — exactly the time that it takes to form one somite. Movement of the pulses does not depend on cell displacement or on propagation of an activating signal, and each cell seems to have its own functional clock³⁵.

Because of its intrinsic free-running property, the cell division cycle must also be considered as a molecular clock. Most eukaryotic cells in culture undergo mitosis (nuclear division) with a periodicity of roughly 24 hours. Is this just coincidence, or were cells sensitive to light–dark cycles millions of years ago? If they were, what we study today as the cell cycle could represent a vestigial circadian rhythm.

PS.-C.

tain a PAS domain¹⁷. Overall it is an acidic protein, containing a probable nuclear-localization signal and an acidic transcriptional activation domain. But Tim, as Per, lacks an obvious DNA-binding domain. So, although there is convincing evidence that Per and Tim act at the level of transcription, it is still not clear how they do it. One crucial step seems to be entry of the Per–Tim dimer into the nucleus^{19,20}. This step, as well as association of the two proteins, has been shown to be regulated by light^{21–23}, although it is not yet known how. Thus, similar to the way in which the production of Frq is controlled by temperature^{11,12}, there is a direct link between the molecular gears of the clock and the physiological modulation of rhythmicity.

Association of the clock molecules seems to be critical for these proteins to function. Is association itself a clock mechanism? In that case, it may be modulated by protein modifications and, therefore, under the control of signalling pathways.

Mammalian oscillators

Although a mammalian equivalent of *tim* has yet to be identified, two mammalian (mouse) *per* genes have been found, encoding the Per1 and Per2 proteins. As in *Drosophila*, Per1 and Per2 both contain a PAS domain^{24–26}. This is classically split into the PAS-A and PAS-B subdomains, and these are divided by a stretch of amino acids that differs between Per1 and Per2. The mouse PAS domains are strikingly similar to that found

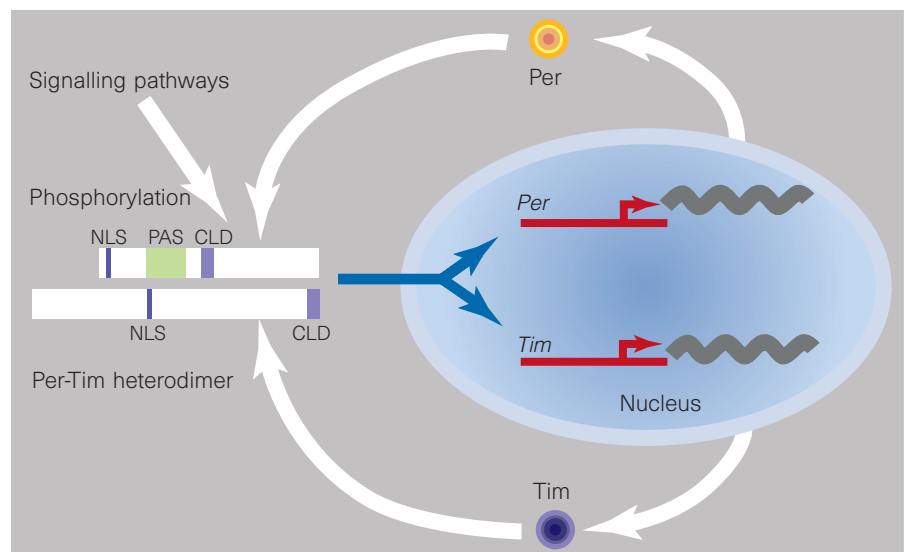


Figure 2 The interdependent regulation of Per and Tim in the fruitfly. These two proteins are co-expressed in neurons, where they interact. This association allows them to enter the nucleus where they regulate transcription of the *per* and *tim* genes. Per contains a PAS structural domain, which is essential for the interaction with Tim. Although Tim does not contain a PAS domain, it contains a putative acidic transcriptional activation domain, which may be involved in regulating the expression of *per* and *tim*. Both Per and Tim contain a nuclear-localization signal (NLS) and a cytoplasmic-localization domain (CLD), which is responsible for retention in the cytoplasm of one protein in the absence of its partner²⁰. Mammalian Per proteins also contain a basic helix–loop–helix (bHLH) domain, which could direct binding to DNA.

in *Drosophila* Per, suggesting that the mouse and fly Per proteins could share structural and functional properties (Table 1). However, mouse Per1 and Per2 also contain a bHLH domain²⁶. So mammalian Per

proteins could act in a different way to the *Drosophila* protein, because of their potential to bind to (as-yet unidentified) DNA regulatory sites.

The mouse *per* genes are expressed in the SCN^{24–27}. In constant darkness they both show circadian patterns of expression which overlap, but are four hours out of phase. The genes differ in their response to light at night, however — whereas expression of *per1* is rapidly induced, *per2* is unresponsive²⁶. Thus, *per1* may be the pacemaker component that responds to light, and may mediate coupling to it (Fig. 1).

A clock factor that could interact with the mouse Per molecules is the product of the *Clock* gene — mice that carry a single mutation in the Clock protein have a dramatically altered circadian rhythmicity²⁸. Clock contains a PAS domain (Table 1), and the PAS domains in Clock, and mouse and fly Per, have similar sequences. Clock also contains a bHLH domain²⁸. So, it is possible that Per modulates the activity of Clock. Or, perhaps Clock associates with other co-factors or with (as-yet unidentified) Per- and Clock-like molecules. Because mutation of *Clock* induces dramatic changes in mouse locomotion, it must constitute a real clock element²⁸. The same cannot be said for *per1* and *per2* until mutation studies have been done.

The expression of circadian genes is not limited to neuronal tissue. For example, as well as being expressed in the SCN (where it shows a dramatic circadian oscillation²⁹), the rat albumin D-element-binding protein (DBP) is a transcription factor that is

Table 1 Features of clock transcription factors

Name	Structural domains	Species	Expression	Function
Per	PAS	<i>Drosophila melanogaster</i>	Hundreds of neuronal and glial cells. Only eight neurons in silk moth.	Regulates clock. Associates with Tim.
Tim	NLS Acidic activation domain?	<i>Drosophila melanogaster</i>	Co-expressed with Per.	Regulates clock. Associates with Per.
Per1	PAS, bHLH	Mouse	Various tissues. Oscillates in SCN. Light-pulse inducible.	Not established.
Per2	PAS, bHLH	Mouse	Various tissues. Oscillates in SCN. Not light inducible.	Not established.
Clock	PAS, bHLH Q-rich (Activation domain?)	Mouse	Various tissues. Not reported oscillating.	Essential for mouse clock function.
Frq	NLS, H-turn-H Activation domain?	<i>Neurospora crassa</i>	---	Essential for clock function.
Fos	bZIP Activation domain	Vertebrates	Various tissues. Light-pulse inducible in SCN.	Output. Associates with Jun constituting AP-1.
Jun-B	bZIP	Vertebrates	Various tissues. Light-pulse inducible in SCN.	Output. Associates with Fos. Repressor of Jun.
ICER	bZIP	Vertebrates	Various tissues. Cyclic AMP inducible. Oscillates in pineal gland.	Output. Repressor. Regulates melatonin synthesis.
DBP	bZIP, PAR Activation domain	Vertebrates	Circadian expression in liver and SCN.	Output. Possibly regulates cyclic cholesterol synthesis.

Per, Tim, Clock and Frq are part of the clock mechanism, whereas Fos, Jun-B, ICER and DBP are implicated in the control of output signals.

bHLH, basic helix–loop–helix; NLS, nuclear-localization signal; Q-rich, glutamine-rich; H-turn-H, helix–turn–helix; bZIP, basic region (involved in DNA binding) and leucine zipper (which directs dimerization); PAR, proline- and acid-rich domain (possibly involved in transcriptional activation).

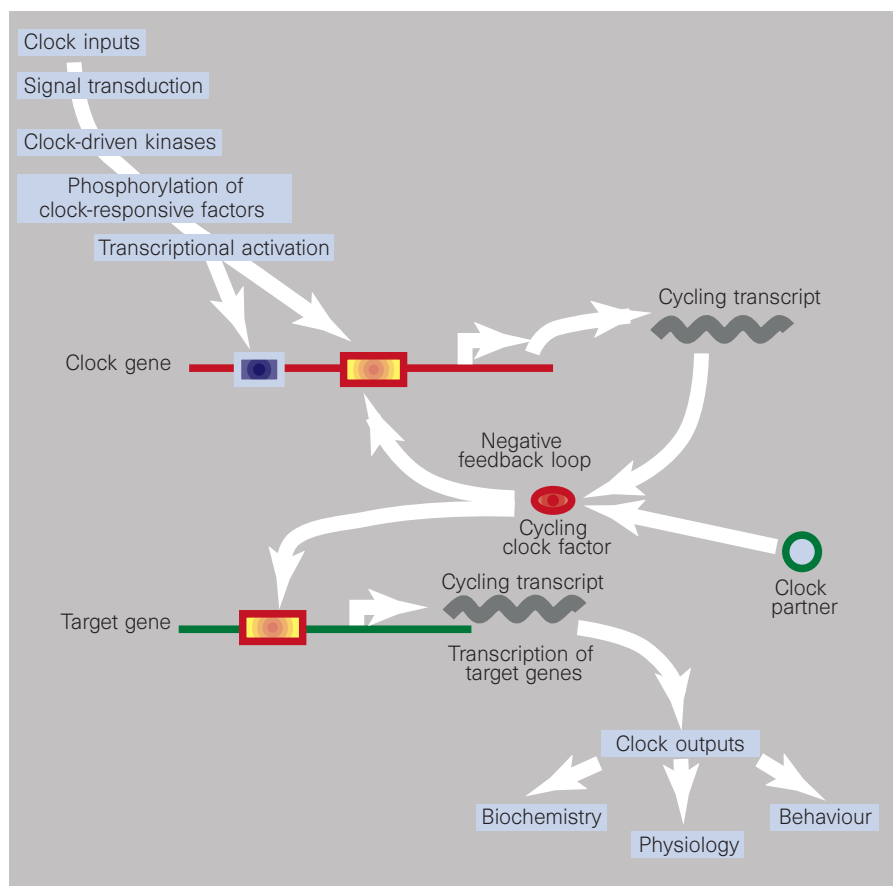


Figure 3 Feedback, autoregulatory loops are central for clock function. Clock factors may bind to their own promoters, leading to negative feedback (whereby transcription of the clock gene is turned off) or positive feedback (whereby transcription is turned up). The clock factor may also modulate the promoter of another (target) gene, leading to expression of that protein in an oscillating fashion. Additional proteins (clock partners) may be required to obtain the feedback.

enriched in the liver. The cyclic expression of DBP follows a circadian rhythm, and it is regulated at the transcriptional level²⁹. It may be influenced by oscillating levels of corticosteroid hormones, which are produced in the adrenal cortex. Interestingly, DBP regulates the expression of cholesterol 7 α -hydroxylase, levels of which fluctuate during the day, yet it does not seem to regulate transcription of albumin, which is not cyclic. Thus, a transcription factor that undergoes a circadian modulation may be involved in the control of liver physiology and, possibly, of cholesterol balance.

Autoregulatory loops

The Frq protein in *Neurospora*^{10–12} and Per¹⁷ in *Drosophila* act through autoregulatory loops — that is, they rely on feedback to regulate their own expression. In the fly, the Per autoregulatory loop is connected to a loop made by Tim, and this link regulates transport of the Per–Tim complex to the nucleus¹⁷ (Fig. 2). Such feedback loops are probably a hallmark of rhythmicity³⁰ (Fig. 3).

A remarkable example of this is given by the cyclic-AMP-responsive element modulator (CREM), which is expressed cyclically in the pineal gland in response to clock sig-

nals from the SCN⁴. Levels of CREM are elevated at night but absent during the day and, as with *per* and *frq*, this cyclic expression is regulated at the transcriptional level. The oscillating messenger RNA encodes a very small factor termed the inducible cAMP early repressor (ICER). This acts as a very powerful transcriptional repressor, and it can switch off expression of its own gene in a cyclical fashion^{4,31}.

Why is this transcriptional repressor regulated in a circadian manner, and what does this have to do with controlling physiology of the pineal gland? One link could be the rhythmicity of melatonin synthesis — ICER regulates the expression of serotonin-N-acetyltransferase (NAT), which is the limiting enzyme in the pathway for the synthesis of melatonin. So, the rhythmic expression of ICER, and its feedback loop, may help to switch off the production of melatonin by cyclically repressing transcription of the NAT gene³². These findings implicate CREM in shaping the oscillatory synthesis of hormones, and highlight the fact that a transcriptional loop is critical in physiological control⁴.

Such autoregulatory loops seem to be key features of many clock systems. But how

similar are the other molecular features of the clock in different species? Per, for example, seems to be structurally and functionally similar in the fruitfly and mammals, yet it may work in a different way. And does *Drosophila* have a Clock protein? Do mammals have a Tim? If so, will the expression of, and interactions between, these proteins be the same in different species? The answers to these questions will improve our understanding of how the clock works, by providing insights into the molecular gears that constitute its mechanism. In any case, today we know much more than only ten years ago, when the direct involvement of transcription factors had not been considered. Indeed, the relatively recent effects of molecular biology in the field have provoked stunning advances — an optimistic prophecy of the discoveries yet to come. □

Paolo Sassone-Corsi is at the Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM, Université Louis Pasteur, BP 163, 67404 Illkirch, Strasbourg, France.

e-mail: paolosc@igbmc.u-strasbg.fr

- Pittendrigh, C. S. *Annu. Rev. Physiol.* **55**, 17–54 (1993).
- Takahashi, J. S. *Annu. Rev. Neurosci.* **18**, 531–553 (1995).
- Klein, D. C., Moore, R. Y. & Reppert, S. M. *Suprachiasmatic Nucleus — The Mind's Clock* (Oxford Univ. Press, New York, 1991).
- Foulkes, N. S., Borjigin, J., Snyder, S. H. & Sassone-Corsi, P. *Trends Neurosci.* **20**, 487–492 (1997).
- Ralph, M. R., Foster, R. G. & Menaker, M. *Science* **247**, 975–978 (1990).
- Tosini, G. & Menaker, M. *Science* **272**, 419–421 (1996).
- Plautz, J. D., Kaneko, M., Hall, J. C. & Kay, S. A. *Science* **278**, 1632–1635 (1997).
- Ralph, M. R. & Menaker, M. *Science* **241**, 1225–1227 (1988).
- Morris, M. E. *et al. Science* **279**, 1544–1547 (1998).
- Luo, C., Loros, J. J. & Dunlap, J. C. *EMBO J.* **17**, 1228–1235 (1998).
- Liu, Y., Garceau, N. Y., Loros, J. J. & Dunlap, J. C. *Cell* **89**, 477–486 (1997).
- Garceau, N. Y., Liu, Y., Loros, J. J. & Dunlap, J. C. *Cell* **89**, 469–476 (1997).
- Konopka, R. J. & Benzer, S. *Proc. Natl Acad. Sci. USA* **68**, 2112–2116 (1971).
- Hall, J. C. *Trends Neurosci.* **18**, 230–240 (1995).
- Hardin, P., Hall, J. C. & Rosbash, M. *Proc. Natl Acad. Sci. USA* **89**, 11711–11715 (1992).
- Mittag, M., Lee, D. H. & Hastings, J. W. *Proc. Natl Acad. Sci. USA* **91**, 5257–5261 (1994).
- Reppert, S. M. & Sauman, I. *Neuron* **15**, 983–986 (1995).
- Huang, Z. J., Edery, I. & Rosbash, M. *Nature* **364**, 259–262 (1993).
- Vosshall, L. B. *et al. Science* **263**, 1606–1609 (1994).
- Saez, L. & Young, M. W. *Neuron* **17**, 911–920 (1996).
- Zeng, H., Qian, Z., Myers, M. & Rosbash, M. *Nature* **380**, 129–135 (1996).
- Myers, M. P., Wager-Smith, K., Rothenfluh-Milfiker, A. & Young, M. W. *Science* **271**, 1736–1740 (1996).
- Hunter-Ensor, M., Ousley, A. & Seghal, A. *Cell* **84**, 677–685 (1996).
- Tei, H. *et al. Nature* **389**, 512–516 (1997).
- Sun, Z. S. *et al. Cell* **90**, 1003–1011 (1997).
- Albrecht, U., Sun, Z. S., Eichele, G. & Lee, C. C. *Cell* **91**, 1055–1064 (1997).
- Shearman, L. P., Zylka, M. J., Weaver, D. R., Kolakowski, L. F. & Reppert, S. M. *Neuron* **19**, 1261–1269 (1997).
- King, D. P. *et al. Cell* **89**, 641–653 (1997).
- Lopez-Molina, L., Conquet, F., Dubois-Dauphin, M. & Schibler, U. *EMBO J.* **16**, 6762–6771 (1997).
- Sassone-Corsi, P. *Cell* **78**, 361–364 (1994).
- Molina, C. A., Foulkes, N. S., Lalli, E. & Sassone-Corsi, P. *Cell* **75**, 875–886 (1993).
- Foulkes, N. S., Borjigin, J., Snyder, S. H. & Sassone-Corsi, P. *Proc. Natl Acad. Sci. USA* **93**, 14140–14145 (1996).
- Grainger, R. M. & Gordon, J. B. *Proc. Natl Acad. Sci. USA* **86**, 1900–1904 (1989).
- Temple, S. & Raff, M. *Cell* **44**, 773–779 (1986).
- Palmeirim, I., Henrique, D., Ish-Horowitz, D. & Pourquie, O. *Cell* **91**, 639–648 (1997).



*The Artist Takes Aim on a Target for Optical Imitation: one of Albrecht Dürer's "perspective machines", published in his *Introduction to Measurement* in 1538.*

Kemp's conclusions

What has emerged in this series, despite the core of commonalities linking old masters, classic modernists and living artists, is a clear change in shape for the relationships between art and science across the three areas.

Martin Kemp

Having reached the end of *Nature's Art and Science* series, the time seems right for a summing up. The essays, each centred on a single artefact, were written to a three-part cycle of 'old master' (from the Renaissance to mid-nineteenth century), 'classic modernist' (from the first half of this century) and living artist. Against this pattern were plotted other variables: artistic content; the type of science involved; the style of art-science relationship; artistic medium; geographical spread; and tone of artistic voice. Not all these variables have been respected as much as others — unsurprisingly, given my own uneven knowledge and the personal factors involved in making choices.

Notable vitality

There has been an obvious leaning towards British artists in the essays on living practitioners, reflecting my own contacts, but there need be no apology for highlighting British practice at this time of notable vitality. I have dealt only with western art — so there were no Islamic tile patterns or Chinese nature painting. For the most part I have selected artists to whom I feel a commitment, Salvador Dali being the most notable exception (*Nature* 391, 27; 1998). Hence, no M. C. Escher, the favourite of mathematicians and the ingenious designer of tessellations and spatial conundrums, whom I regard as heavy-handed and aesthetically laboured.

The guiding principle has been to seek what I am calling 'structural intuitions', rather than to describe instances of the

influence of science on art, or (occasionally) vice-versa. By 'structural intuitions' I mean those elements in our perception and modelling of the seen world (including what we believe lies *behind* appearance) that are the common province of anyone who indulges in the critical exercise of our visual faculties in the most searching manner — whether artist or scientist. The 'structural' element relates both to the patterns that can be extracted from nature at various levels of complexity and to the structures that operate in our perception, cognition and visualization. If this sounds alarmingly circular, I should say that I believe firmly in the non-arbitrary relationship between how nature works and the apparatus with which we have been provided — by nature and by nurture — to make functional sense of the sensory chaos of the world.

Human 'structural intuitions' as expressed visually seem to me increasingly to show an enduring core of commonalities across ages and cultures, however different the modes and vehicles of cultural expression. One manifestation of this is the recurrence of decorative motifs in very different civilizations, such as spiral formations, polygonal patterns and diverse symmetries.

That being said, the art historian's job is largely concerned with the search for what is special about a specific cultural moment, emphasizing differences in style, mode of visualization, function, patronage, reception and so on. During the course of the series, what emerged with unexpected clarity — for the author at least — was the nature of the changing shape to the relationships between art and science across the three areas

of old master, modernist and present. This shape was not predetermined, and the three-part cycle worked against the perception of works chronologically. The shape may be described, telegraphically, as analytical description, abstraction and process.

Analytical description refers to a form of representation in which aspects of appearance are remade — literally re-presented — on the basis of an intuitive or intellectual understanding of the nature of what is being seen, how it is seen and how it may be depicted in such a way as to convey 'information' to an attuned viewer. The tools for such remaking include the projective system of linear perspective, the rational description of light effects through shadow and the modulation of colour (manipulation of tone, hue and saturation), and the structural bases of natural form (such as human anatomy or geology).

Leonardo da Vinci, with whom we started (*Nature* 389, 799; 1997), worked with all these tools, and no masters of naturalism could afford to neglect any of the optical means, however selectively they might concentrate on other areas of interest. Given the prominence of empirical analysis in so many sciences during the 'Scientific Revolution', it is not surprising that some artists were able to play an active role in the dialogue between various types of seeing and knowing. They contributed not only in the obvious areas of illustration but also in the more searching evocation of the causes of natural effects — whether the perceptual explorations of Jan Vermeer (*Nature* 392, 27; 1998) or the fiery emissions from the inner Earth in Joseph Wright's volcanoes (*Nature* 391, 645; 1998).

A signal characteristic of modern science is a high degree of remove and abstraction from the sensory parameters of our normal experience, thanks to equipment used to see and often to generate emissions inaccessible to our eyes — X-rays, infra-red, thermal radiation, sonar, electrons and other sub-atomic particles. Spectacular devices enable us to 'see' forms and forces at scales of minuteness and immensity unimaginable with conventional microscopes and telescopes. Particles have etched the febrile geometry of their kamikaze lives across the spaces of cloud and bubble chambers, bearing witness to a realm in which the classical laws of physics are inapplicable. At the same time, modern artists are striving to forge a realm of artistic reality separate from that of the eyewitness and of conventional naturalism. They are seeking an aesthetic autonomy that obeys its own set of unfathomable rules — whether these are seen as residing in the mind or in the greater forces of the Universe, or both in concert.

It is best not to become obsessed with the question of influence, and of seeking demonstrable ways in which ideas from the new physics percolated into educated consciousness (which they undoubtedly did). Rather, we should see art, in its own right, as exploring submerged worlds of mind and matter that had only been implied under the skin of older naturalisms, now that the burden of naturalism had shifted to the photographic media. The desired autonomy of non-figurative art may be manifested in forms hostile to science or in conscious harmony with scientific explorations. In either case, the idea that truth (to whatever) is located in modes of representation that are not dependent upon and may even contradict our commonsense notions of realism places modern art in a reciprocal relationship to the counter-intuitive revelations of modern scientific theories such as relativity and quantum mechanics.

What Lewis Wolpert has called the "unnatural nature of science" came to the fore at the very time that artists such as Picasso were forging an 'unnatural' manner of representation, in which the picture was seen working according to rules specific to the nature of 'pictorial reality' rather than traditional naturalism. The advent of Freudian psychology, as reflected in the collages of Max Ernst (*Nature* 392, 137; 1998), also posited forms of unconscious reality that were quite different from the logical relationships between things we expect in the external world.

How and why these unnatural formulas came to be shared by arts and sciences in the early years of this century are questions that are best for the moment entered through the individual gateways of artists such as Max Ernst, Umberto Boccioni (*Nature* 391, 751; 1998), Naum Gabo (*Nature* 389, 919; 1997),

Max Bill (*Nature* 390, 239; 1997) and Josef Albers (*Nature* 390, 451; 1997), who arrived at comparable degrees of remove from standard visual experience down quite different paths. It would be nice to propose a neat general theory with some confidence, but I suspect the factors behind unnatural art were so complex and variable, and stood in such diverse positions with respect to scientific knowledge, that the best the historian can do is to chart the individual 'mutations' and to search for the conditions that let them thrive.

Dissolving boundaries

The characterization of contemporary art in terms of process is clearly dependent on my choice of artists. Given the plurality of current practice, a different selection might emphasize such tendencies as the need for instant media impact, a tiresomely incestuous self-reference to art and artiness, a resort to found objects, or a dissolving of the fixed boundaries inherent in any definition of 'Art'.

But, at this disadvantageously close viewpoint, it seems to me that one of the factors — perhaps *the* factor — that distinguishes the artistic cutting edge in the last two decades has been process, not so much its description in the manner of Wright's volcanoes but rather the use of a process that is allowed to determine the final configuration of the work — if there is indeed any 'final configuration' rather than an animated image of William Latham's type (*Nature* 391, 849; 1998). A programme is initiated, with chosen parameters, but the end result is not predetermined.

This lack of determinism, even when the means are apparently simple as with 'Callan's canyons' (*Nature* 390, 565; 1997) or Glen Onwin's vats of brine (*Nature* 391, 543; 1998), has clear affinities with non-deterministic chaos and self-organised criticality — those fashionable kinds of iterative computation that have reclaimed the visual dimension for advanced mathematics. And the self-similar properties of fractals stand in a suggestive relationship to the scale-less qualities that frequently characterize artworks which use process as the way of manipulating the chosen medium. But are the relationships more than suggestive?

Jonathan Callan was as surprised as I to be told by Adrian Webster that his "dust-scapes" conformed to Voronoi cells and exhibited "affinities with a recent model of the architecture of the Universe on very large scales" (*Nature* 391, 430; 1998). Clearly, there is no influence at work here. Rather, what I believe is happening is that one of the dominant tenors of an age

obsessed with process — through computer programs, notions of market forces, the remorseless march of 'selfish genes', big bangs and black holes, the transmission of information through cyberspace, and so on — is being addressed by artists, in the same way that artists have always created images framed by the leading paradigms of their society.

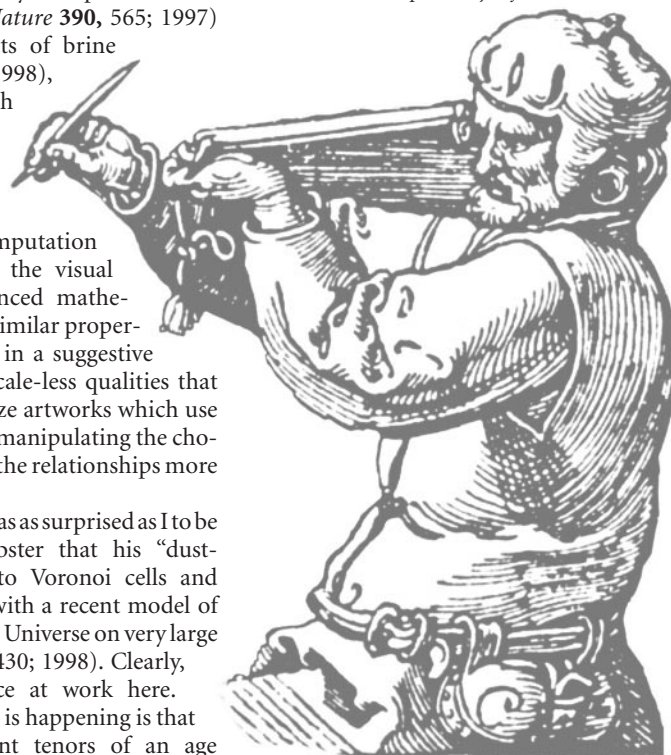
Shared intuitions

The artist and scientist both live within, and play active roles in constructing, human mental and physical landscapes. That they should share 'structural intuitions' is less surprising than inevitable. What is surprising and wonderful is how these intuitions have manifested themselves in the works of innovative artists and scientists in culturally apposite ways.

Where next? Our plan is to publish a companion series on visual images from scientific cultures, ranging from models of big molecules to the physiognomy of Dr Jekyll's Mr Hyde, and from the periodic table to a science laboratory. I will be exploring the 'look' of scientific things, to see whether the style of scientific artefacts is no less integral to the communication of content and meaning than style in a 'work of art'. The 'sci-artefacts' will be drawn from the technical and popular histories of their sciences and serve as transmitters within various worlds of communication — the wider areas of science and learning, institutions and funders, and various kinds of public. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk



Painters centre one eye in portraits

The importance of the centre of the canvas has long been appreciated in art¹, as has the importance of the eyes in revealing the personality of subjects of portraits. The centre of symmetry of the face is often discussed in art analysis¹⁻⁵ and might be expected to be used as an explicit principle of composition by artists trained according to such analysis. However, I have found that portraits painted throughout the past 600 years adhere to a different compositional principle not discussed in the literature: one eye is consistently centred horizontally in the canvas.

Fig. 1 illustrates the degree to which an eye tends to be set near the horizontal centre in six classic portraits selected with the heads in a variety of poses, with no attempt at a scientific sampling. Artists often set one eye on the centre line even when the portrait departs from the classic three-quarter pose.

To quantify the relation between eye position and the canvas frame, the horizontal positions of the eyes were measured in portraits from the past 600 years — including many from the 20th century — by all 265 portrait painters represented in a variety of published summary sources. Portraits selected were the first occurring in each source by every artist that were hand-drawn (oil paintings, watercolors, drawings or engravings), and that depicted only one person, from above the waist, with both eyes visible. Figure 2a, b and c shows the distributions obtained. For comparison, the positions of the centres of the mouths were also measured (Fig. 2c).

I defined the most-centred eye of a portrait as the one closest to the vertical centre line. If eyes were positioned according to the centre of symmetry of the two eyes in relation to the vertical axis, both eyes will be about the same distance from the axis and the choice of eye will make little difference to the result. Conversely, if the head is positioned randomly around the centre vertical, choice of the closer eye as the one for analysis will narrow the distribution somewhat, but by no more than a factor of $\sqrt{2}$, the standard deviation of the minimum of two samples from a gaussian distribution.

The artists surveyed placed one eye in a narrow distribution peaking at the horizontal centre (s.d. = $\pm 5.6\%$ of the frame width). A similarly narrow distribution (s.d. = $\pm 5.5\%$) was exhibited by one-third of the sample in which the faces were the most frontal, estimated at less than 10° head-turn on the basis of the displacement of the tip of the nose from the halfway point between the eyes. The narrowness of this distribution shows that the eye centring is not simply a result of the head-turn common in portraits.

Conversely, the position of the mid-point



Figure 1 Eye-centring in classic portraits. The portraits are reproduced at the full width and arbitrarily cropped at the bottom; the white line runs down the centre. The examples (left to right within each layer) are by Rogier van der Weyden (c. 1460) Sandro Botticelli (c. 1480), Leonardo da Vinci (1505), Titian (Tiziano Vecellio; 1512), Peter-Paul Rubens (1622) and Rembrandt van Rijn (1659).

between the two eyes forms a bimodal distribution (Fig. 2b) as expected if one or other eye were being centred (s.d. = $\pm 12.0\%$). This mid-point distribution is significantly different from a gaussian distribution at $P < 0.01$ on the χ^2 test, whereas the best-centred eye distribution is well-fitted by a gaussian distribution at $P > 0.1$. Again, the most-frontal third of the paintings exhibited a distribution for the mid-points that was significantly broader (s.d. = $\pm 11.1\%$, $P < 0.01$) than that for the most-centred eye, and not significantly narrower than for the whole sample. So even in frontal portraits, where the face is shown in its most symmetrical view, it is not typically placed symmetrically in the frame.

Other features of the face seem to be in-

accurately centred. The horizontal position of the mouth, for example, was spread across the frame (Fig. 2c), with a distribution about three times wider than that of the best-centred eye (s.d. = $\pm 15\%$, significantly wider than $1.414 \times 5.6\%$, $P < 0.01$).

However, one exception to the eye-centring principle is for the side view of the head. In the smaller sample of side views from the same sources⁴⁻¹¹, the eye positions were scattered widely throughout the frame (s.d. = $\pm 25\%$, significantly wider than $1.414 \times 5.6\%$, $P < 0.01$). It is unclear what aspect of side views changes the rules of composition. Perhaps, now the subject's attention is seen as being directed away from the viewer, the principle becomes the

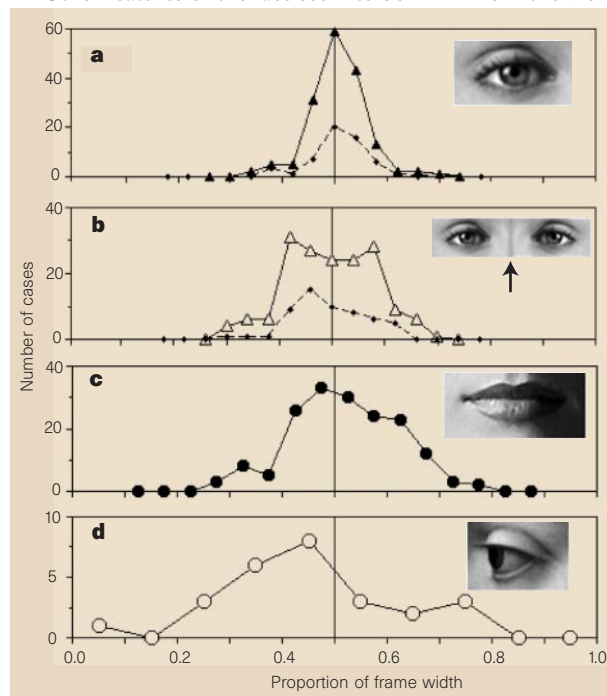


Figure 2 Relation between eye or mouth position and canvas frame in portraits painted over the past 600 years. **a**, Histogram of the middle of the opening for the most-centred eye. Triangles, all 265 portraits; diamonds, one-third of the sample in which the faces were most frontal. **b**, Distribution of the mean position of the mid-point between the eye centres. Triangles, all 265 portraits; diamonds, distribution of the one-third of the sample in which the faces were most frontal. **c**, Horizontal mouth positions in portraits. **d**, Horizontal eye positions in the 23 profiles (portraits with only one eye visible) from the same sources.

centring of the head in the frame because the head dominates the composition.

Classical texts on composition lack any mention that the eyes should be positioned relative to the frame of the picture, but instead typically emphasise the placement of centres of mass in the frame, or relative to the vanishing point in cases of central perspective^{1–5, 12–18}. If art analysis omits eye-centring as a compositional principle, its manifestation throughout the centuries and varieties of artistic styles must be essentially unconscious.

It is interesting to compare the placement accuracy in portraits with that in a psychophysical study of error in the placement of elements within a frame¹⁹. Reproduction of the position of a single dot was accurate to about $\pm 2\%$ of frame width, while accuracy fell to about $\pm 5\%$ for the placement of four or more dots simultaneously. It seems that the unconscious (or unexpressed) placement of the eye in portraits is nearly as accurate as the attentive performance of those focusing on positioning as their sole task.

My analysis shows that explicit compositional principles are implemented with an unbiased accuracy of $\pm 5\%$ over the past six centuries. This precision results from perceptual processes that seem to be unexpressed by the artists themselves, suggesting that hidden principles are operating in our aesthetic judgements, and perhaps in many realms beyond portraiture.

Christopher W. Tyler

Smith-Kettlewell Eye Research Institute,
2232 Webster Street, San Francisco,
California 94115, USA
e-mail: cwt@skivs.ski.org

1. Arnheim, R. *The Power of the Center Ch.* I–IV (Univ. Calif. Press, Berkeley, 1988).
2. Arnheim, R. *Art and Visual Perception: A Psychology of the Creative Eye* 36–41; 89–90 (Univ. Calif. Press, Berkeley, 1954).
3. Jaxtheimer, B. W. *How to Paint and Draw* 143–146 (Thames & Hudson, London, 1982).
4. Edwards, B. *Drawing on the Right Side of the Brain* 168 (Tarcher/Putnam, New York, 1989).
5. Goldwater, R. & Treves, M. *Artists on Art, from the XIV to the XX Century* 51 (Pantheon, New York, 1945).
6. Schneider, N. *The Art of the Portrait: Masterpieces of European Portrait Painting 1420–1670* (Taschen, Cologne, 1994).
7. Petersen, K. & Wilson, J. J. *Women Artists: Recognition and Reappraisal from the Early Middle Ages to the Twentieth Century* (Women's Press, London, 1978).
8. The Smithsonian Staff. *The Smithsonian Experience* (Smithsonian Instit., Washington DC, 1977).
9. Coates, T. *Creating a Self-Portrait* (Mitchell Beasley, London, 1989).
10. Wilson, S. *Holbein to Hockney: A History of British Art* (Bodley Head, London, 1979).
11. Robinson, J. M. *Windsor Castle Official Guide* (Joseph/Penguin, London, 1995).
12. Alberti, L. B. *On Painting* (1435) trans. Spencer, J. R. (Yale Univ. Press, New Haven, 1966).
13. Ruskin, J. *The Laws of Fesole (1877–1879)* (Allworth Press, New York, 1996).
14. Henri, R. *The Art Spirit* (Lippincott, Philadelphia, 1923).
15. Gombrich, E. H. *Art and Illusion: A Study in the Psychology of Pictorial Representation* (Princeton Univ. Press, Princeton, 1969).
16. Kubovy, M. *The Psychology of Perspective and Renaissance Art* (Cambridge Univ. Press, Cambridge, Massachusetts, 1986).
17. Dunn, C. *Conversations in Paint* (Workman, New York, 1995).
18. Kent, S. *Composition* (Dorling Kindersley, London, 1995).
19. Igel, A. & Harvey, L. O. *Gestalt Theory*, 13, 210–231 (1991).

Ball court design dates back 3,400 years

Excavations at the archaeological site of Paso de la Amada, in the Soconusco region of Pacific coastal Chiapas, Mexico, have uncovered an earthen ball court dating to approximately 1400 BC (uncalibrated), which is at least five centuries older than any previously excavated ballcourt in Mesoamerica¹. Moreover, this discovery reveals that the design of ball courts dates back 3,400 years.

At the time of the Spanish Conquest, ball courts were elaborate open-air masonry structures with two parallel platforms and a central alley where players competed to pass a rubber ball through a wall-mounted hoop. Ball games and their associated rituals were more than simple sporting events. They were integral to the political, religious and social life of most ancient Mesoamerican societies². The importance of the ball game is reflected in the frequent placement of ball courts around ceremonial plazas in city centres and their close association with temples, palaces and other administrative buildings³.

The ball game itself has been thought to pre-date the appearance of formal ball courts. Archaeological evidence of the ball game includes ceramic representations of ball players in their regalia^{4,5}, as well as the water-logged remains of latex rubber balls⁶ dating to about 1250 BC.

The Paso de la Amada discovery demonstrates that formal ball courts have long been an important component of the ball game complex and have been present from

the very inception of settled village life.

The initial discovery of the ball court at Paso de la Amada was accidental. In the course of the 1995 excavations at the site, we began probing Mound 7 in search of remains of Early Formative Period (1600–900 BC) households (Fig. 1a). As this mound was the largest of 54 visible mounds at the site, measuring 110 m by 50 m by 2 m high, we anticipated finding evidence of residential structures similar to those encountered in other areas of the site. Unexpectedly, a trench bisecting the mound revealed two parallel earthen platforms, with benches 2.5 m wide and 35 cm high flanking a central alley 80 m long (Fig. 1b). These architectural elements are unique to ball courts and are similar to those found in later sites.

Further excavation of the mound allowed us to determine its stratigraphy and document the following sequence (Fig. 1b): (1) pre-construction use of the location during the Barra phase (1550–1400 BC); (2) initial ball court construction early in the Locona phase (1400–1250 BC); (3) amplification of the ball court later in the same phase; (4) gradual sedimentation of the alley during the Locona and Ocós (1250–1100 BC) phases; and (5) the abandonment of the court and erosional infilling of the alley during the Chelra phase (1100–1000 BC).

Two radiocarbon samples bracketed the construction dates of the ball court. The court was built some time after the Barra phase surface (1490 $\pm$ 50 BC, B-82233) on which it sits and before the accumulation of sediments that washed down from the lateral platforms onto the alley floor late in the Locona phase (1270 $\pm$ 60 BC, B-82234).

The discovery and dating of this ball court indicates first, that large-scale ball courts, requiring significant amounts of labour, were in use much earlier than previously thought; and second, that ball court form was conserved with few modifications until the Spanish Conquest. We suggest that this ball court could have been part of a network of similar courts in the Soconusco region during the Early Formative period.

Although we know very little about this early version of the ball game, we can infer from its later variants that a strong element of conflict and competition — both real and ritualized — was involved in playing the game. A network of ball courts may have provided villagers in the Soconusco region with a means of intervillage competition, while simultaneously helping to maintain community solidarity. If this were the case, we would expect to find more early ball courts in the region, but so far no others have been found.

We also suggest that emerging elites within villages may have sponsored the construction of ball courts in order to enhance their status and prestige both locally and regionally². This implies that the earliest ball courts

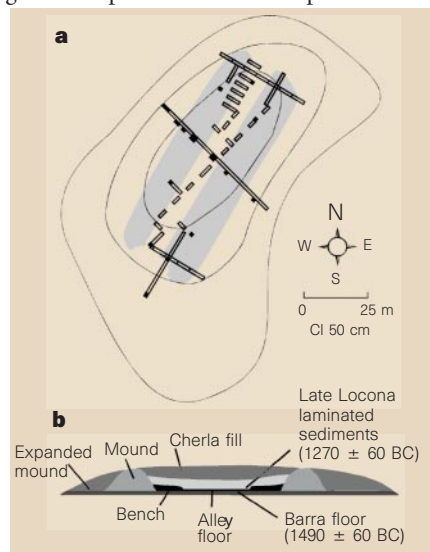


Figure 1 The ball court at Paso de la Amada. **a**, Plan view showing the outline of the ball court and the 1995 excavations at Mound 7. CI, contour interval. **b**, Cross-section of Mound 7 showing features of the ball court and locations of the radiocarbon samples. Radiocarbon dates and archaeological phases are presented in uncalibrated radiocarbon years.

should be located in sites with evidence of status differences among community members. At Paso de la Amada, we have mounting evidence for both high- and low-status households that are contemporary with the ball court⁷⁻¹⁰. The location and excavation of more early ball courts should help to establish their role in the emergence of social and political complexity.

Warren D. Hill, Michael Blake

University of British Columbia,
Department of Anthropology and Sociology,
6303 NW Marine Drive, Vancouver,
British Columbia V6T 1Z1, Canada
e-mail: whill@anso.ubc.ca

John E. Clark

Brigham Young University,
Department of Anthropology, 950 SWKT, Provo,
Utah 84602, USA

1. Agrinier, P. in *The Mesoamerican Ballgame*, (eds Scarborough, V. L. & Wilcox, D. R.) 175–176 (Univ. Arizona Press, Tucson, 1991).
2. Fox, J. G. *Current Anthropol.* **37**, 483–509 (1996).
3. Scarborough, V. L. & Wilcox, D. R. (eds) *The Mesoamerican Ballgame* (Univ. Arizona Press, Tucson, 1991).
4. Oliveros, A. in *El Juego de Pelota en Mesoamérica: Raíces y Supervivencia* (ed. Uriarte, M. T.) 39–51 (MT Siglo XXI editores, Mexico).
5. Ekholm, S. in *The Mesoamerican Ballgame* (eds Scarborough, V. L. & Wilcox, D. R.) 241–250 (Univ. Arizona Press, Tucson, 1991).
6. Ortiz, P. & del Carmen, M. R. in *Los Olmecas en Mesoamérica* (ed. Clark, J. E.) 69–92 (Citibank, Mexico, 1994).
7. Clark, J. E. in *The Formation of Complex Society in Southeastern Mesoamerica* (ed. Fowler, W. R. Jr) 13–26 (CRC, Boca Raton, 1991).
8. Blake, M. in *The Formation of Complex Society in Southeastern Mesoamerica* (ed. Fowler, W. R. Jr) 27–46 (CRC, Boca Raton, 1991).
9. Lesure, R. L. *Lat. Amer. Antiquity* **8**, 217–236 (1997).
10. Clark, J. E. & Blake, M. in *Factional Competition and Political Development in the New World* (eds Brumfiel, E. M. & Fox, J. W.) 17–30 (Cambridge Univ. Press, Cambridge, UK, 1994).

New ice outdoes related nets in smallest-ring size

In a recent Letter, Lobban, Finney and Kuhs¹ reported a remarkable new metastable form of ice (ice XII) in the field of stability of ice V and presented an accurate structure determination. The authors state correctly that the topology of the phase is unlike that of any of the known phases of ice. They also state that it “contains a mixture of five- and seven-membered rings of water molecules” which “is interesting, although not unexpected”. This second statement is not completely true; the truth is more interesting and, I believe, unexpected, as is agreed by the authors of ref. 1 (personal communication). The structure contains only seven- and eight-membered rings and is the first example of a 4-connected net of this type.

The 4-connected nets of interest in crystal chemistry can all be realized as 4-coordinated sphere packings, in which the sphere centres correspond to the vertices of the net and the sphere contacts correspond

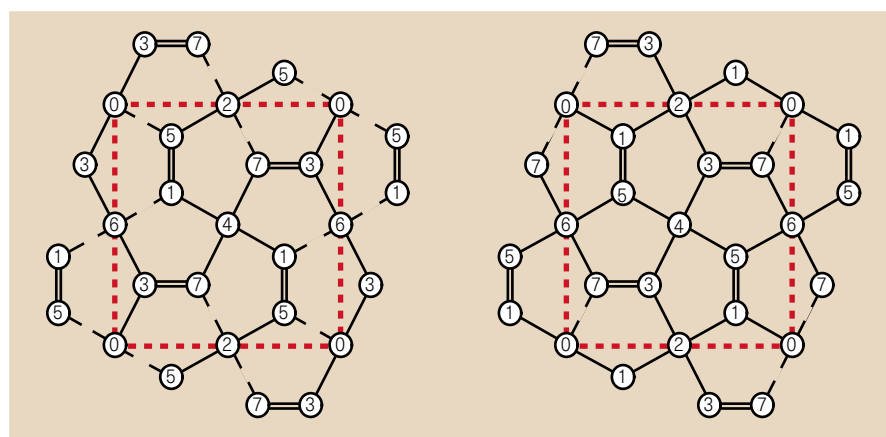


Figure 1 Nets projected down c with numbers giving elevations in multiples of $c/8$. Left, the ice XII net. Links between vertices are shown as lines, pairs of links up and down as double lines, and links between vertices in adjoining cells as broken lines (for example, a line between vertex at elevation 6 and one at elevation 1 is really a link between vertices at elevations -2 and 1, or at 6 and 9). The dashed square outlines a unit cell. Right, a second net with the same symmetry as that of ice XII but of lower density.

to edges. This ensures that each vertex has exactly four equidistant nearest neighbours. In practice, a further restriction is useful²; in the conformation of maximum volume subject to the constraint of equal edges, the structure should remain 4-coordinated.

The net of the new ice structure fits these criteria, and I give coordinates for maximum volume and equal (unit) edge length. The symmetry is $I\bar{4}2d$ with $a = 2.988$, $c = 1.424$, $c/a = 0.477$. The vertices are of two kinds: A in $4a, 0,0,0 \dots$ and B in $8d, x,1/4,1/8 \dots$ with $x = 0.3675$. This is close to the conformation ($c/a = 0.485$, $x = 0.3643$) found in the real material¹.

The vertex symbols³, which give the number and kind of smallest rings contained in each angle, are for A: $7_2 \cdot 7_2 \cdot 7_2 \cdot 8_4 \cdot 8_4$ and for B: $7_3 \cdot 7_3 \cdot 7_3 \cdot 8_4 \cdot 8_4$. It turns out that every angle containing 7-rings also contains two 8-rings, and the total ring count for the unit cell is sixteen 7-rings and twenty-four 8-rings.

Figure 1 (left) shows the net projected down c . It should be clear that there are no 5-rings. I find this structure remarkable because every other known 4-connected net of interest in crystal chemistry contains smallest rings that are 6-rings or smaller. This is true for all the 130 or so zeolite structures; indeed, it is true for all 400 or so individual vertices of these structures.⁴

It is also true for all my collection of more than 200 uninodal and binodal nets, most of which occur in one or more contexts in crystal chemistry. From a study of such structures, it was concluded⁵ that “a tetrahedral framework whose smallest rings are 7-membered or more will probably not be possible”.

A net has been described³ with only 7-rings, but this is an exceptional structure that cannot be realized as a 4-coordinated sphere packing and with one angle that lacks

rings. It is not likely to serve as a framework for structures such as ices and aluminosilicates.

The new structure (in the maximum-volume form given above) is one of the densest 4-connected nets in known crystal structures (I exclude intergrowths of two or more nets such as are found in some dense forms of ice). For unit edge length, the number of vertices per unit volume is $r = 0.944$. This might be compared with $r = 0.845$ for coesite (the densest 4-connected silica net) and $r = 0.750$ for quartz.

The most nearly comparable structure is that of ice IV (ref. 6), which also occurs metastably in the ice V stability field and for which I find $r = 0.971$ in the maximum-volume form. The vertex symbols for the two vertices (which occur in the ratio 1:3) in this structure are $6_2 \cdot 8_2 \cdot 6_2 \cdot 8_2 \cdot 6_2$ and $6 \cdot 6_2 \cdot 6 \cdot 8_3 \cdot 6 \cdot 8_3$ (this structure also contains 10-membered rings).

Another net (Fig. 1, right) that has the same symmetry as that of ice XII but is of lower density ($r = 0.669$) does contain 5- and 7-rings and appears in simulated annealing experiments⁷ as a silicon net in SiO_2 . For this net (again with unit edges), $a = 3.193$, $c = 1.760$, and the vertices are $5 \cdot 5 \cdot 5 \cdot 5 \cdot 8_2 \cdot 8_2$ at $0,0,0$, and $5 \cdot 5 \cdot 5 \cdot 7 \cdot 7 \cdot 7$ at $0.1556, 1/4, 1/8$.

Michael O’Keeffe

Department of Chemistry, Arizona State University,
Tempe, Arizona 85287, USA

1. Lobban, C., Finney, J. L. & Kuhs, W. F. *Nature* **391**, 268–270 (1998).
2. O’Keeffe, M. & Brese, N. E. *Acta Crystallogr. B* **48**, 663–669 (1992).
3. O’Keeffe, M. & Hyde, B. G. *Crystal Structures I. Patterns and Symmetry* (Mineralogical Society of America, Washington DC, 1996).
4. O’Keeffe, M. & Hyde, S. T. *Zeolites* **19**, 370–374 (1997).
5. Brunner, G. O. *Zeitschrift für Kristallographie* **156**, 296–303 (1981).
6. Engelhardt, H. & Kamb, B. J. *Chem. Phys.* **75**, 5887–5899 (1981).
7. Gibbs, G. V., Boisen, M. B., O’Keeffe, M. & Bartelmehs, K. L. (in the press).

Designs on evolutionary invention

Cats' Paws and Catapults: Mechanical Worlds of Nature and People

by Steven Vogel

Norton: 1998. Pp. 382. \$27.50.

To be published in the UK in October.

R. McNeill Alexander

The earliest patent for barbed wire, of 1868, says the invention was designed to give wire the character of a thorn hedge. The same message is given by the name of the manufacturer, the Thorn Wire Hedge Company. The inventor of modern chain-saw teeth had studied the jaws of a wood-boring beetle, and the inventor of Velcro was inspired by burrs. These inventions are well documented as having imitated nature, but it would be hard to find many more to add to the list. We often want to do things that nature does well, but we seldom copy nature's methods.

Engineers have looked to nature for ideas more often than they have imitated it directly. Otto Lilienthal wrote a book entitled *Bird Flight as the Basis for Aviation*, and his pioneering gliders of the 1890s had wings of cloth stretched between radiating spars, very much like the wings of bats which are stiffened by long radiating fingers. But the Wright brothers and their successors built wings with a rectangular grid of spars, unlike any animal wing, and no successful aircraft has been powered by flapping wings. In his new book comparing human and natural engineering, Steven Vogel argues that flapping wings are fine for birds that fly relatively slowly, but could not work well for large, fast aircraft.

Our preference for rectangular grids, straight lines and square corners contrasts with nature's preference for curves, especially spirals. Our rectangular products stack neatly (imagine what supermarket shelves would be like if cereal packets were shaped like snail shells), and they lend themselves to constructions built from many identical modules. Spiral structures such as snail shells give a different advantage: they can be extended without changing their shape. If you extend your garage without demolishing any of the existing walls you will get a longer garage that is relatively narrower, but a snail adds to the edge of its shell and enlarges it while keeping the same proportions. Cylinders, however, are found frequently both in nature and in engineering, often (in both cases) reinforced by helically wound fibres. Examples include pressure hoses and parasitic worms.

Different priorities, in natural and human engineering, explain why we make rectangular boxes while nature makes spiral shells. There are also major differences in materials and power sources. Nature cannot

extract metals from their ores, which generally requires high temperatures and large quantities of energy. It cannot use internal combustion engines (again, high temperatures are involved), and its lack of good electrical conductors makes electric power impractical. In some respects, however, nature performs much better than human engineering. Bone is a composite material that exploits the same principles as glass fibre, but its fibres are orientated to match the loads it has to bear far more subtly than we can manage with manufactured materials, and it adapts automatically when loading patterns change.

Our workshops and kitchens are full of lever systems designed to amplify forces — pliers, wire-cutters, nutcrackers and can-openers. In contrast, as Vogel points out, animal lever systems usually amplify movement at the expense of force — muscles shorten by small amounts to move our hands and feet through much larger distances. That is a

consequence of the properties of our muscles, which can lengthen and shorten only by small fractions of their length, but it seems paradoxical that, to exert the large forces needed to crack a nut, we have to use force-amplifying nutcrackers to counteract the force-reducing lever system of our own skeleton.

Vogel's book has been written to be enjoyed, with the same clarity as his previous books. It is full of ideas and well-explained principles that will bring new understanding of everyday things both to non-scientists and scientists alike. It covers some of the same ground as Michael French's *Invention and Evolution* (Cambridge, 1988) and James Gordon's *The Science of Structure and Materials* (Scientific American Books, 1988), but having read those fine books is no reason for avoiding this one, which offers plenty of new insights. □

R. McNeill Alexander is at the School of Biology, University of Leeds, Leeds LS2 9JT, UK.

Listen out for the death rattle

This is as close as most people would want to get to a San Lucas rattlesnake (*Crotalus ruber lucasensis*). But for those who want to know

more, Chris Mattison's easy-to-read *Rattler! A Natural History of Rattlesnakes* is now out in paperback (Blandford, \$19.95, £10.99).



People power

The Art of Moral Protest: Culture, Biography, and Creativity in Social Movements

by James M. Jasper

University of Chicago Press: 1998. Pp. 514.

\$35, £27.95

Steven Yearley

The classical social theorists of a century ago anticipated that the industrialized capitalist societies would offer new but coherent identities, and that these would lie at the heart of political protest in the modern period. Foremost among these was Karl Marx. He and, in various ways, his followers argued that the manifold forms of identity on offer — based on craft, region, nationality or empire allegiance and even gender — would gradually resolve themselves into class. Although the Marxist interpretation showed this analytical move in its clearest light, within the social sciences there was a widespread assumption that individuals would come to take on one (or a few) out of a small range of identities.

The politics of abortion or animal rights were never supposed to become truly important. Subsequently, many commentators have drawn attention to the emergence of 'new' social movements which, increasingly through the twentieth century, have offered new collective identities and multiple bases for protest. For many strict Marxists and other social scientists who had assumed that there were only limited alternatives, the dramatic increase in the range of political identities on offer has been a puzzle and a challenge to deeply held commitments. James M. Jasper attempts to bring together the vast literature on these protest movements and aims to offer a systematic framework for analysing them.

It is easy to agree with Jasper's estimation of the importance of these protest movements. They have commanded attention in several important ways: through their success in attracting committed members; through their legislative and policy achievements; and through the innovative forms of political activity they have stimulated. For example, in the United Kingdom alone the environmental movement has between four million and six million members, as represented by people who voluntarily commit money and time to participation in environmental activities. In the summer of 1997, the Royal Society for the Protection of Birds — hardly the most glamorous of organizations — celebrated the recruitment of its millionth member.

The total membership of UK environmental groups now approaches the number of people who are members of trades unions and far exceeds the membership of the



Captive audience: Greenpeace activists demonstrating in Moscow against the planned French nuclear tests at Mururoa atoll in 1995.

Labour Party. In relation to legislative achievements, the civil rights movement — in the United States, in South Africa and even in Northern Ireland — has secured for previously excluded minorities the benefits of formal equivalence before the law in matters of voting and has improved equality in employment and pay. In relation to political innovation, the women's movement has pioneered new forms of protest and has politicized issues previously regarded as quite beyond the realms of the political. These issues have recently centred on domestic violence, but formerly included a whole range of interpersonal issues now commonly seen as regulated by 'political correctness'.

Jasper provides a lengthy account of existing ways of trying to understand social movements. For example, a leading US approach focuses on social movement organizations (such as Greenpeace) and on the competition for members, funds and other resources. He points out that if the term 'resources' is allowed to swell too much, it becomes useless as an explanation. In any case, he observes, some protest movements appear to be initiated and run by individuals. And some protests (such as many animal rights campaigns) appear to recruit people through literature and shocking photographs, which can easily be produced by individuals, rather than through the networking activities of sophisticated organizations. Furthermore, he stresses the role of emotional responses in the setting up and development of protest movements. Too often, in his view, social scientists have overlooked this idiosyncratic and apparently irrational component of political protest.

In place of an emphasis on one kind of explanatory factor, Jasper elaborates a four-fold model of the 'dimensions' of protest. He acknowledges that resources are important,

but attributes comparable importance to the strategic decisions made by protesters, the biographical background of particular protesters and the cultures that protesters share. By 'cultures' he means both the beliefs and the emotional outlook of protesters.

Jasper uses a range of examples of political protest — from boycotts in nineteenth-century Ireland to anti-nuclear campaigning in northern California — to make a reasonably persuasive case for the importance of these four dimensions. As his title implies, he is consistently concerned with the 'artfulness' — by which he means the creativity and innovativeness — of protest.

But his approach seems to suffer from two shortcomings. First, he does not have any particularly precise theories about the operation of his four 'dimensions'. So all the case-study material he cites seems compatible with his approach, yet the cases could hardly be said to test the approach in any meaningful sense.

Second, apart from some references to ideas recently publicized by the eminent UK sociologist Anthony Giddens, Jasper says little with any authority about changes in the character of social movements. Some commentators have argued that the end of the twentieth century is a particularly fertile time for the growth of social movements; according to some Italian authors we are entering 'tribal' times.

Whether because large sections of Western populations no longer expect scarcities and therefore concentrate on diverse, so-called post-material concerns, or because a decline in public trust in expertise leads to protest over successive new technologies, or because new social movements are prone to divisiveness, social movements are apparently becoming more numerous. Jasper does not use his four dimensions to explain (or

even to deny) this trend. By the end of the book, one does not know whether to expect our society to become more 'tribal' or not. Jasper has made a contribution to analytical thinking about the structure of protest movements, but he has not resolved the social scientific puzzle of the multiplication of identities in the advanced industrial world. □

Steven Yearley is in the Department of Sociology, University of York, York YO10 5DD, UK.

Current concerns

Joseph Henry: The Rise of an American Scientist

by Albert E. Moyer

Smithsonian Institution Press: 1998. Pp. 348. \$45, £34.95

Jacqueline Reynolds and Charles Tanford

One of the main research problems for nineteenth century physicists was the interconversion of electricity and magnetism, and not only from a purely theoretical viewpoint. The practical applications that were to emerge — the dynamo, the electromagnetic telegraph and the transformer — influenced the lives of everyone, not just the researcher in his laboratory.

Of the many familiar names associated with this subject, first and foremost is the incomparable Michael Faraday (1791–1867) of the Royal Institution in London, a man from a poor family, largely self-educated, mathematically illiterate but clear-thinking and imaginative. Hermann von Helmholtz said of him: "Faraday performed in his brain the work of a great mathematician without using a single mathematical formula."

However, during the same period, in the fledgling United States, another scientist paralleled Faraday's work in electromagnetism but, although eulogized by the US population at his death, is now nearly forgotten. Joseph Henry (1797–1878), both in his background and his career, was uncannily similar to Faraday.

Henry was born into a Scots Presbyterian family of little means, and what formal education he managed to acquire at the small Albany Academy in New York State was paid for by dint of his own labours. He held strong religious beliefs and through most of his research career was well removed from politics of any sort. During his years as a teacher at Albany Academy, and later as professor of natural philosophy at Princeton (then the University of New Jersey), Henry built electromagnets of great strength, generated electric currents from changing magnetic fields, discovered mutual induction and self-induction, and built a working electromagnetic telegraph, all activities engaged in by his counterpart at the Royal Institution during the same period.

But Henry differed from Faraday in many particulars. He was not blessed with a mentor of the quality of Sir Humphry Davy, nor did he have the financial support that Faraday received from the Royal Institution. As a scientist in the distant former colonies, Henry was isolated from the kind of interaction that Faraday had with his numerous friends and colleagues in the European scientific community. Most important, whereas Faraday's maxim was "work, finish, publish", Henry was slow to finish and publish his research, perhaps because of pressures of his heavy teaching responsibility, but certainly also in no small part owing to his own inertia. And herein lies the principal reason for his relative lack of credit for what were often simultaneous discoveries.

Albert E. Moyer has produced a fascinating private and public picture of Henry, based in large measure on the archives held at the Smithsonian Institution in Washington DC. The great wealth of material available — including laboratory notebooks, students' comments on his lectures, and private letters — is skilfully interwoven with descriptions of Henry's experimental activities, family life and public contacts. Moyer draws an intimate picture of a Janus-like character.

The public face was described by Spencer Fullerton Baird, his assistant and successor as director of the Smithsonian, in a biographical sketch written for the *Encyclopaedia Britannica*: "He was a man of varied culture, of large breadth and liberality of views, of generous impulses, of great gentleness and courtesy of manner, combined with equal firmness of purpose and energy of action." Many of his contemporaries thought Henry to be modest and self-effacing.

Privately, however, a different picture emerges of a man often far from modest in his own estimation and tortured by his perceived lack of appropriate recognition by the scientific community; of an individual with little ability to judge his own work objectively. Henry wrote an anonymous biography of himself for the *Princeton Review* which dwelt at length on his many accomplishments, including a list of his 22 most significant research projects and discoveries, including the first electromagnets powerful enough to lift more than a ton, the first continuously operating machine based on an electromagnet, the development of the concept and apparatus essential for electromagnetic telegraphy and the discovery of self-induction.

Another anonymous article appeared in the *Encyclopedia Americana* reviewing the advances in electricity and magnetism before 1847 in which he elevated his own work to a foremost position. "After the discovery by Dr. Faraday of the foregoing principles of galvanic induction, the most important additions to this branch of electricity have been made by Prof. Henry."

Moyer's biography regrettably deals only sparsely with the last 30 years of Henry's life, after his appointment in 1846 as secretary of the newly founded Smithsonian Institution, a position he did not seek but was honoured to accept when offered. Henry was in effect the sole director, and the job required all his energy and time. Research into the nature of electromagnetism was abandoned, and there is no indication in Moyer's book as to whether or not he followed the developments in this field. It would be intriguing to know his response to the great treatise of James Clerk Maxwell, which was published five years before Henry's death. Did he even know about it? Would he have understood it? Probably not, we suspect. □

Jacqueline Reynolds and Charles Tanford are at Turlwood, Back Lane, Easingwold, West Yorkshire YO6 3BG, UK.

Fossil dichotomy

Palaeoecology: Ecosystems, Environments, and Evolution

by P. Brenchley and D. Harper

Chapman and Hall: 1997. Pp. 432. £29.99, \$44.95

Brian R. Rosen

The arrival of the first wholly new palaeoecological text in English for 15 years is a noteworthy event. Most palaeontologists would probably claim an interest in palaeoecology but there have been surprisingly few texts to guide them, and these provide very different accounts.

As pioneering authors in the 1960s, R. F. Hecker gave us practicalities and D. V. Ager showed that ancient natural history could be fun. A. J. Boucot showed that palaeoecology could be useful, J. W. Valentine showed that it could be seriously evolutionary, and W. S. McKerron led a glass-bottomed boat cruise through ancient communities. J. R. Dodd and R. J. Stanton gave us rigour but at the price of a rationale — taxonomic uniformitarianism — that was too confined for dealing with long-extinct organisms. So an outsider might well think that palaeoecology is an eclectic residue of 'fascinating things you can do with the fossil record,' once you have dealt with hard-core taxonomy and evolutionary systematics.

One problem lies with palaeoecology's conflicting origins. Is it an Earth science or a life science? The mode of life of ancient organisms helps geologists to interpret past environments (Earth science), but geological evidence for past environments helps to reveal the mode of life of ancient organisms (life science). Or, as Dodd and Stanton relate in a telling anecdote: "We asked [a sedimentologist] if he could identify the depositional environment. He replied that... if we could tell him which [environments] the fossils

In retrospect by Mikhail Mina

The Gospel of Afranius: The Holy History as an Object of a Detective Inquiry

by Kirill Es'kov

Published in Russian by the author

rasn@glasnet.ru

1995

During the Soviet period in Russia, 'militant atheism' was a cornerstone of the official ideology. Scientists were ordered to be atheists, the result predictably being that even those who did not believe were in sympathy with the believers. It seemed that the fields of religion and science were different and no problem would arise if there was no trespassing from either side. Most scientists avoided participating in official antireligious campaigns, and religious activists had neither desire nor opportunity to start the brawls.

The situation changed with the fall of the communist regime. Russian scientists began to find all kinds of missionaries and mystics grazing in their kitchen garden, some of them migrants from outside, some mutated from the local stock. Now the former owners feel uneasy. It still seems unethical (and hardly possible) to kick the trespassers out, but it is also humiliating to endure the invasion silently. Such is the historical context within which *The Gospel of Afranius* should be assessed.

Kirill Es'kov is a specialist in the palaeontology and zoogeography of spiders, and a senior scientist at the Palaeontological Institute of the Russian Academy of Sciences in Moscow. The appearance of *The Gospel of Afranius* was stimulated by *The Resurrection Factor* by Josh McDowell (Thomas, 1981). McDowell maintains that he analysed all the possible materialistic explanations of events following the crucifixion of Christ as they are described in the New Testament and found none satisfactory. In his view, it is enough to conclude that here we are faced with the direct intervention of God.

Es'kov defines his own position clearly and precisely. He writes: "As for religion, I am an agnostic like many of my colleagues — naturalists. For me it has always been an axiom that there is not and cannot be a proof of God's existence in the sphere of mind. Since Protestant McDowell has rejected Tertullianus' honest 'Credo, quia absurdum' and with his own hands desacralized the text of the Gospel, he



deliberately got involved in a rather risky game on the field of the opponent. Unable to withstand the temptation I have accepted his challenge."

After that Es'kov demonstrates what a specialist accustomed to

analysing fragmentary and not very reliable data can do even in an area outside his normal domain. He does it brilliantly in the first part of the book, discovering a lot of 'logical possibilities' overlooked by the opponent even though, playing fair, he accepts 'presumption of honesty', excluding any version that implies fraud committed by Christ or the apostles.

The second part of the book contains a story written, the author assures us, by Afranius, the chief of Pontius Pilate's Secret Service, a person well known to Russian readers as a character in a popular novel, *The Master and Margarita* by Mikhail Bulgakov. The story describes an operation of the Secret Service that used Christ without his knowledge. The witty narration is excellently stylized as a modern 'spy novel' and perfectly dovetails into the *Gospel's* scenario.

The story explains how skilful professionals could stage a 'resurrection' using substitutes after Christ died on the cross and manage to convince the apostles that they had seen nobody but their resuscitated rabbi. The author says: "You wanted a materialistic explanation, didn't you? All right, here you are!" The humour is sometimes biting, but never insulting.

The manuscript was offered to several publishers but none wanted to spoil its relations with the Russian Orthodox Church, so the author had to publish the book himself. It has already won admirers. In 1997, it won the Grand Prix at the Festival of Science Fiction Authors in Odessa. There is reason to think that *The Gospel of Afranius* would find many interested readers if it were published in English.

Mikhail Mina is at the Laboratory of Postnatal Ontogeny, Institute of Developmental Biology, Russian Academy of Sciences, Vavilov St. 26, Moscow 117334, Russia.

indicated, he could identify the sedimentary environment!" For marine palaeoecologists at least, the Earth science tendency has surely been uppermost, whereas D. R. Lawrence, in a seemingly long-forgotten appraisal published in 1971, reclaimed it as a life science.

This dichotomy is not all. As Lawrence also argued, palaeoecologists have relied too much on applying recent ecology to the past, "yet this methodology proves nothing since working methods presuppose static

natural histories through time". Instead, he challenges us to concentrate on the environmental indications of fossil-bearing sediments, and on the inherent evidence of the fossils themselves. In other words, for palaeoecology to go forward it has to be structured around rigorous methodologies for reconstructing the mode of life of ancient organisms, independently of simplistic uniformitarianism. I believe this is vital if we are to translate mere assemblages

of fossils more soundly into ancient communities, and thereby allow palaeoecology to realize its most exciting potential: unravelling the history of communities.

So there is a real dilemma in writing a palaeoecological text. Should it keep faith with its eclectic but problematic traditions, and dutifully cover everything studied in its name, or should it be based on a consistent rationale? Brenchley and Harper's contents list reveals a pragmatic solution that undoubtedly does justice to the subject as it has been studied. They give a rich range of palaeoecological case histories, organized under straightforward headings, some as Earth science (environmental indicators), some as life science (populations and communities, trace fossils), some as essential background (environmental controls, taphonomy), and just one not strictly palaeoecological at all (palaeobiogeography). The scale ranges from detailed considerations of the morphology of individual organisms to the global history of the biosphere.

The younger parts of the fossil record, and the palaeoecology of carbonate environments considering the latter's biotic richness and role in carbon burial and climate change, are not well covered, while the intriguing ecology of modular and social organisms is overlooked. And some of the references seem a little old.

Nevertheless, students and teachers will find this by far the most satisfactory palaeoecological text to date. Yet, nearly 30 years on, Lawrence's challenge still stands. □

Brian R. Rosen is in the Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 10 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 1996 and issued at least four separate numbers by the end of May 1998.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language is English.
- Deadline for submission is 5 June.

Please send at least four different issues (the first, the most recent and any two others) of each eligible title, together with full details of subscription rates, to: Peter Tallack, *Nature*, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567. e-mail: p.tallack@nature.com

Allosteric effects of DNA on transcriptional regulators

Jeffrey A. Lefstin & Keith R. Yamamoto

Selective gene transcription is mediated in part by regulatory proteins that bind to DNA response elements. These regulatory proteins receive global information from signal-transduction events. But transcriptional regulators may also be modified in an allosteric manner by response elements themselves to generate the pattern of regulation that is appropriate to an individual gene.

The end-point of many, if not most, signal-transduction pathways is modification of a transcriptional regulator by covalent, extracellular or polypeptide ligands. Here we discuss recent progress suggesting that response elements can be ligands for regulators, acting as allosteric effectors that determine the regulators' conformation. A given regulator may therefore activate transcription in the context of one gene, repress transcription in another, and bind but exert no regulatory effect in a third. Thus, we propose that response elements contain information that is interpreted by bound regulatory factors.

Implicit in this notion are extensions of two conventional views of transcriptional regulation: first, although response elements position regulators close to the promoters that they serve, they may also inform the bound factor about the mode of regulation to be conferred at that site; and second, although regulatory factors exhibit modular organization, such that isolated domains display subsets of the activities of the intact factors, those activities may also interact in the intact factors, yielding distinct patterns of regulation.

Hence DNA acts as an allosteric ligand whose binding alters the regulator's affinity for other ligands, such as coactivators or corepressors. If the regulator assumes distinct conformations on binding to different genomic sites, then the DNA sequence of the response element can specify which surfaces are available to contact target factors, and thereby determine how the protein regulates transcription at that site (Fig. 1a, b). This mechanism is versatile and adaptable; for example, transcriptional regulators could be 'tethered' through protein-protein contacts with DNA-bound factors, without themselves contacting DNA. In such cases, the bound protein would serve as the allosteric effector (Fig. 1c).

Modularity of transcriptional regulators?

Allosteric control of regulatory-factor activity by DNA requires a specific functional connection between the DNA-binding and transcriptional-regulatory surfaces. However, eukaryotic transcriptional regulators are commonly considered to be modular proteins whose regulatory regions and DNA-binding domains (DBDs) operate independently¹. This viewpoint derived from the general success of 'domain-swap' experiments, in which an 'activation domain' or 'repression domain' demonstrates intrinsic activity when delivered to a promoter by a heterologous DBD, such as that of the yeast GAL4 protein.

However, the demonstration that DNA-binding domains and regulatory domains can operate independently when uncoupled need not imply that they have no functional interaction in their native context. By analogy, tyrosine kinases of the Src family contain separable domains that, when separated, can interact with various other proteins; in the intact kinase, however, these domains display intramolecular associations that regulate kinase activity².

Some activation domains are not autonomous

Some transcriptional-regulatory domains do not behave as autonomous modules. Instead, they are sensitive to the identity of their attached DBD, or to their DNA-binding state. These findings are not

common, perhaps because standard strategies for identification of regulatory domains typically depend on sufficiency tests in which the candidate region is fused to the DBD of GAL4. Domains that do not function autonomously will not be detected unless comparative analyses with the native DBD are conducted. For example, the transcriptional regulator USF2 contains a conserved domain, the USF-specific region, that activates transcription when attached to the USF2 DBD. But the USF-specific region is completely inactive, or even repressive, when fused to the DBD of GAL4 (ref. 3).

In other cases, native DBDs exert inhibitory effects on cognate transcriptional-activation functions. Serum-response factor (SRF) and ATF2, transcriptional regulators of the MADS and bZIP families, respectively, contain separable activation domains that activate transcription when tethered to a promoter by the GAL4 DBD. Unexpectedly, the full-length proteins remain silent when tethered by GAL4. However, deletion of the MADS or bZIP DBD restores transcriptional activation by the chimaeras^{4,5}. In the native protein, the activation domain may be silenced by the DBD, perhaps until it interacts with a response element.

Oct-4 contains a domain that activates transcription in many cell types when tethered to a promoter by the GAL4 DBD. However, in the context of the native Oct-4 POU-type DBD, this activation domain functions only in certain cell types⁶. The Oct-4 activation

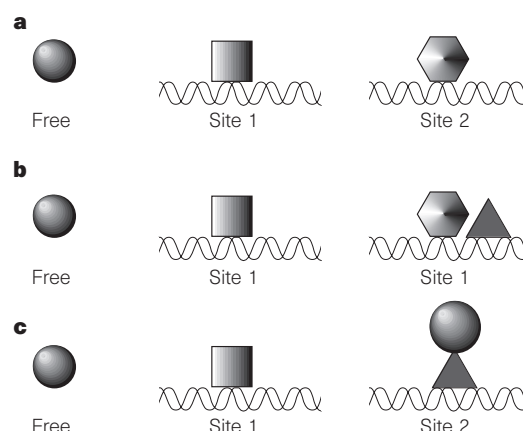


Figure 1 Different effects of DNA response elements on regulator function. **a**, A regulator (shaded objects) adopts different conformations when directly bound to different specific DNA sequences. **b**, A regulator's conformation is determined by a combination of DNA sequence and the influence of a nearby protein (triangle). **c**, A transcriptional regulator, which binds directly to site 1, is recruited to site 2 solely by protein-protein contact with a site-2-specific factor (triangle). With any of these modes, the choice of surfaces presented by the regulator might depend on the oligomerization state of the protein, conformational changes within a multifunctional DBD, allosteric effects transmitted by intramolecular contacts, or a combination of these mechanisms.

Table 1 DNA-induced conformational changes in eukaryotic transcriptional regulators

Protein	DNA-binding-domain motif	Conformational change	Method	Ref.
GCN4	bZIP	Helix induction/stabilization	CD, NMR, X-ray	9, 55, 56
C/EBP	bZIP	Proteolytic resistance	Protease digestion	10, 57
c-Jun, c-Fos	bZIP	Helix induction/stabilization	CD, NMR, FTIR	11, 58
MCM1	MADS	Proteolytic sensitivity	Protease digestion	14
Ets-1	wHTH	Proteolytic sensitivity, helix destabilization	Protease digestion, CD	7
c-Myb	HTH	Helix induction/stabilization, proteolytic resistance, hydrophobic-core rearrangement	NMR, CD, protease digestion	59–62
MASH-1	bHLH	Helix induction/stabilization	CD	8
Pax-6	Paired	Helix induction/stabilization	CD	12
NF-κB	Rel	Proteolytic resistance, proteolytic sensitivity, altered CD spectrum	Protease digestion, CD	63–65
Steroid receptors	C4 zinc-finger	Helix induction/stabilization, loop reorientation	NMR, X-ray	37–42
TR	C4 zinc-finger	Proteolytic resistance, altered CD spectrum	Protease digestion, CD	13, 66

CD, circular dichroism; FTIR, Fourier-transform infrared spectroscopy.

domain retains its cell-type specificity when fused to the Oct-2 POU DBD, but activates promiscuously when fused to the POU DBD of Pit-1. Other activation domains fused to the Oct-4 or Oct-2 POU domains do not exhibit such cell-type restriction, indicating that the Oct-4 activation domain acquires its specificity through a precise functional interaction with its native DBD.

Activation domains, then, even if they can operate as separable modules, may be functionally integrated with the DBDs of their native proteins. Such integration implies an active role for the DBD. In the traditional view, DBDs serve only to localize modulatory domains to the vicinity of regulated genes. The allosteric hypothesis demands that DBDs adopt distinct conformations at different response elements, modulating transcriptional-regulatory functions in response to the nature of the binding site. In fact, many DBDs change conformation when they bind DNA. In principle, this structural flexibility provides a mechanism for transcriptional regulators to adopt different conformations at different binding sites.

DNA-induced structural changes

Conformational changes upon binding to DNA may be the rule, rather than the exception, in DNA-binding proteins. Moreover, although relatively few proteins have been studied in detail in both free and bound states, DNA-induced changes in secondary and tertiary structure have been described for a range of eukaryotic transcriptional regulators (Table 1).

In some cases, conformational changes are induced in response to either specific or non-specific DNA^{7,8}, whereas in others they are restricted to, or observed at a greater extent in, specific DNA complexes^{9–13}. In a few cases, a protein bound to two different DNA sequences assumes distinct conformations^{13–15}.

Even subtle structural changes can have long-range functional consequences. For example, the DBD of Ets-1 undergoes little or no detectable change in secondary structure in response to DNA^{6,17}, but in the full-length protein DNA binding induces the dissociation and unfolding of helices far removed from the protein–DNA interface^{7,18}. DNA binding can also affect quaternary structure by determining protein oligomerization state or heterodimerization partners.

DNA-binding domains are multifunctional

If conformational changes in DNA-binding domains are to influence the operation of the transcriptional regulator, we would expect DBDs to harbour activities beyond simple DNA recognition. In fact, DBDs typically constitute surfaces for both intramolecular and intermolecular protein–protein contacts. Intramolecular interactions provide a direct mechanism for transducing information from the DBD to regulatory domains. Such intramolecular associations have been observed for the DBDs of ATF2 (ref. 5), Ets-1 (ref. 19),

and the glucocorticoid receptor (J.A.L. and K.R.Y., unpublished observations). In Ets-1 and the glucocorticoid receptor, these intramolecular contacts are released upon DNA binding.

DBDs also mediate intermolecular interactions. Most DNA-binding proteins can form dimers, and the choice of association state or dimerization partner can affect transcriptional regulatory activity. Dimerization or ‘crosstalk’ between different transcription factors is frequently found to involve each factor’s DBD. For example, the communication between steroid receptors and the activator protein-1 (AP-1) complex involves DBDs²⁰. Several cellular and viral modulators of transcription that do not themselves bind DNA operate through association with the DBDs of transcription factors²¹. Basal transcription factors and coactivators have also been reported to contact the DBDs of transcriptional regulators^{22–24}.

Intracellular receptors

As many transcriptional regulators are conformationally responsive to DNA binding, contain DNA-binding domains with multiple functions that could be modulated by these conformational changes, and possess activation domains that are sensitive to the state of the DNA-binding domain, it seems likely that some of their activities will be governed by site-specific allosteric regulation. Although a matter of debate^{5,25}, studies of Oct-1 (refs 26, 27) and SRF^{28,29} imply that certain DNA sequences induce conformational changes in these proteins that govern the presentation of transcriptional-regulatory surfaces. This is most apparent, however, in the intracellular receptor superfamily of regulators.

The retinoic-acid and thyroid-hormone receptors. Members of one major branch of this family, including the receptors for retinoids, thyroid hormones, and vitamin D, generally bind to cognate response elements as heterodimers with the retinoid X receptor (RXR). These response elements usually consist of two repeats of sequences related to AGGTCA, with the spacing between the two motifs determining the choice of RXR partner.

Differentially spaced half-site DNA elements produce distinct transcriptional responses from the retinoic acid receptor (RAR). On direct-repeat motifs with five intervening base pairs (DR5 elements, found in the RAR-β2 gene promoter), the RAR–RXR complex represses transcription in the absence of hormone, due to association with the SMRT corepressor³⁰. Hormone binding induces RAR to release the corepressor and activate transcription. However, on DR1 elements (found in the promoter of the cytosolic retinol-binding-protein II gene), RAR–RXR heterodimers repress transcription in both the absence and presence of hormone³¹. Whereas hormone-dependent recruitment of coactivator proteins is indistinguishable at DR5 and DR1 elements, RAR–RXR heterodimers at DR1 fail to release the N-CoR corepressor on hormone

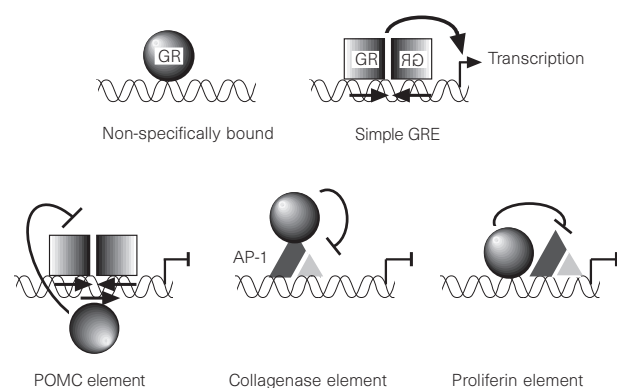


Figure 2 A simple model for control of glucocorticoid-receptor (GR) function by DNA sequence. The DBD of the receptor remains monomeric (top left) unless the dimer interface is induced by a palindromic GRE (top right). In the dimerized conformation, the receptor activates transcription (top right), but unpaired (bottom left) or monomeric (bottom right and centre) GR (circle) is inactive or represses transcription. If recruited to a promoter by particular DNA sequences, protein-protein contacts, or both, unpaired GR can repress the activity of nearby regulators (bottom).

treatment³². Hence, the response of the RAR–RXR complex to retinoids depends on the DNA sequence of the element to which the complex is bound.

Analogous interactions have been observed between the DNA-binding and hormone-binding domains of receptors. Although these domains can function independently when uncoupled, in the intact proteins heterodimerization of RAR and RXR on DNA eliminates hormone binding by the RXR³³. Heterodimerization in solution (or homodimerization of RXR on DNA) has no such effect. This implies that DNA-binding allosterically modifies heterodimeric RXR, altering its hormone-interaction site.

The thyroid-hormone receptor (TR) typically represses transcription in the absence of hormone; hormone binding releases corepressor and promotes activation. However, at response elements associated with the Rous sarcoma virus³⁴, human immunodeficiency virus³⁵, and keratin³⁶ gene promoters, transcription is activated in the absence but not in the presence of hormone. These response elements may allosterically preclude the interaction with corepressor in the absence of hormone, and promote that association when hormone is present. Mutant DR4 elements at which TR–RXR heterodimers bind but fail to activate transcription have been identified¹³; binding of the heterodimer to ‘active’ but not ‘inactive’ DR4 elements is associated with increased proteolytic resistance of TR.

The glucocorticoid receptor. The connection between DNA sequence, protein conformation, multimerization state, and transcriptional activity is perhaps best established in the glucocorticoid receptor (GR). Homodimeric GRs bind to imperfect palindromes of half-sites similar to AGAACA separated by three base pairs; such sites are termed simple hormone-response elements. Binding of the GR to a simple glucocorticoid-response element (GRE) appears to induce two notable conformational changes in the GR DBD: a portion of the dimer interface reorients to contact its opposite number in the second monomer, and a distorted helix which makes both dimer and DNA contacts is stabilized^{37,38}. The distorted helix was detected in some but not all reported structures of GR and oestrogen-receptor DBDs^{39–42}, indicating that it may be metastable and readily influenced by DNA binding or variations in experimental conditions.

Formation of the dimer interface only upon GRE binding accounts for the observation that the GR DBD is monomeric in solution³⁷ but dimerizes cooperatively on simple GREs³⁸. Indeed, the dimerization competence of this receptor correlates with its

ability to activate transcription. Although many natural GREs poorly match the consensus sequence at one of their half-sites, GR will not activate from half-sites alone, or from half-sites with alternate spacings^{43,44}. Moreover, point mutations in the dimer interface severely compromise activation from simple GREs^{45,46}.

Conversely, mutants that constitutively attain the DNA-induced conformation gain GRE-independent transcriptional function. Two independent point mutations in the DBD have been identified⁴⁷ that seem to ‘present’ transcriptional-activation domains under conditions where the wild-type does not. These mutants can activate transcription normally from GREs, but, when overexpressed, interfere *in trans* with transcriptional activation. It was proposed that, in the wild-type receptor, transcriptional-activation surfaces are exposed only when the receptor binds to GREs, but that these mutants constitutively mimic the allosteric effect of DNA. Structural determination by NMR spectroscopy confirmed this prediction: each mutant reorients the dimer interface to the GRE-bound configuration even in the absence of DNA (M.A.A. van Tilborg *et al.*, manuscript in preparation). Disruption of the dimer interface attenuated some but not all phenotypes of these mutants⁴⁷, suggesting that DNA allostery might regulate GR function by both dimerization-dependent and dimerization-independent mechanisms.

Allosteric effects of DNA may help explain how the GR, which contains a strong, autonomously functioning transcriptional-activation domain, can also repress⁴⁸ transcription. Corepressor proteins for GR have not yet been described, but the connection between DNA-induced changes in the dimer interface and transcriptional-activation potential allows us to imagine a simple mechanism: unless the receptor binds the response element as a dimer, a repression function of the receptor is engaged (Fig. 2). For example, at a ‘negative GRE’ from the pro-opiomelanocortin (POMC) gene, a cooperative, trimeric assembly of GRs has been described⁴⁹. Chemical interference data indicate that two of the DBDs are configured as at a simple GRE, whereas the third GR binds, unpaired, to the other face of the DNA, assisted by the binding of the dimeric complex. If the monomeric receptor is repressive, it might silence or antagonize the dimer occupying the opposite face of the DNA.

GR represses transcriptional activation mediated by AP-1 at cognate DNA-binding sites. At the AP-1 site in the collagenase gene, repression occurs without apparent perturbation of AP-1 binding *in vivo*⁵⁰. The *in vitro* association of GR and AP-1 DBDs and the insensitivity of repression to mutation of key residues on the GR DNA-recognition helix^{51–53} indicates that the receptor is ‘tethered’ to AP-1 by protein–protein contacts alone. As GR dimerization is DNA-induced, receptor tethered to AP-1 might be monomeric, with the DBD remaining in the unbound conformation. In fact, mutations in the DBD-dimer interface that eliminate activation from simple GREs do not impair repression⁴⁶, indicating that the receptor may remain monomeric when it is recruited to the collagenase promoter by protein–protein contacts.

The GR also represses AP-1 activation at the composite AP-1 site/GRE associated with the proliferin gene⁵¹. Here both proteins bind to DNA. Although mutations of the GR DBD-dimer interface preclude activation of this receptor by this DNA, they do not interfere with repression⁴⁵. Thus, in three instances of repression, at the POMC, collagenase, and proliferin genes, a monomeric or unpaired GR is implicated as the repressive species (Fig. 2).

Remarkably, missense mutations at a single conserved residue of the GR DBD switch the receptor from repressor to activator at the proliferin and collagenase elements⁵⁴. As this activation also requires integrity of the dimer interface, these mutants further support the notion that the state of the DBD can control the presentation of activation or repression surfaces.

This simple model (Fig. 2) postulates that the dimeric and monomeric states of the GR correspond to the activation and repression states of the molecule, respectively. Strictly speaking,

the conformational state of the DNA-binding domain, rather than the stoichiometry of receptor binding, could direct the choice between activation or repression. One might imagine schemes in which the receptor is recruited as a dimer without engaging the known dimer interface; such species would be repressive according to our model.

Discussion

We have highlighted the ability of transcriptional regulators to adapt their structures, and thereby their functions, to the particular DNA sequences that they recognize. In nature, however, eukaryotic regulators may rarely see their binding sites in isolation. The interaction surfaces presented to a regulator typically include other regulatory proteins and, although response elements can be made up of DNA alone, it is more likely that they comprise a combination of protein and DNA or even (DNA-bound) protein alone. Allosteric control by DNA sequence may merely represent a special case of a general principle: regulatory proteins perform distinct functions depending on the configuration of the macromolecular surface that they recognize.

Although allosteric control of transcriptional-regulatory activities by DNA might seem unexpected, it is, upon reflection, unsurprising. Allosteric mechanisms have evolved in many enzymatic and regulatory proteins, expanding the cell's repertoire of biochemical behaviour without enlarging the genome or establishing new regulatory cascades. In the case of transcriptional regulators, this flexibility allows the organism to use a single protein, which might symbolize cell-positional identity, differentiation state, or the receipt of an extracellular signal, to mediate diverse genetic responses directly. □

Jeffrey A. Lefstin and Keith R. Yamamoto are in the Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California 94143-0450, USA. (e-mail: lefstin@stanford.edu).

1. Ptashne, M. How eukaryotic transcriptional activators work. *Nature* **335**, 683–689 (1988).
2. Nguyen, J. T. & Lim, W. A. How Src exercises self-restraint. *Nature Struct. Biol.* **4**, 256–260 (1997).
3. Luo, X. & Sawadogo, M. Functional domains of the transcription factor USD2: atypical nuclear localization signals and context-dependent transcriptional activation domains. *Mol. Cell. Biol.* **16**, 1367–1375 (1996).
4. Johansen, F. E. & Prywes, R. Identification of transcriptional activation and inhibitory domains in serum response factor (SRF) by using GAL4-SRF constructs. *Mol. Cell. Biol.* **13**, 4640–4647 (1993).
5. Li, X. Y. & Green, M. R. Intramolecular inhibition of activating transcription factor-2 function by its DNA-binding domain. *Genes Dev.* **10**, 517–527 (1996).
6. Brehm, A., Ohbo, K. & Schöler, H. The carboxy-terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain. *Mol. Cell. Biol.* **17**, 154–162 (1997).
7. Petersen, J. M. et al. Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. *Science* **269**, 1866–1869 (1995).
8. Meierhans, D. et al. DNA binding specificity of the basic-helix-loop-helix protein MASH-1. *Biochemistry* **34**, 11026–11036 (1995).
9. Weiss, M. A. et al. Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**, 575–578 (1990).
10. Shuman, J. D., Vinson, C. R. & McKnight, S. L. Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. *Science* **249**, 771–774 (1990).
11. Patel, L., Abate, C. & Curran, T. Altered protein conformation on DNA binding by Fos and Jun. *Nature* **347**, 572–575 (1990).
12. Epstein, J. et al. Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J. Biol. Chem.* **269**, 8355–8361 (1994).
13. Ikeda, M. et al. Different DNA elements can modulate the conformation of thyroid hormone receptor heterodimer and its transcriptional activity. *J. Biol. Chem.* **271**, 23096–23104 (1996).
14. Tan, S. & Richmond, T. J. DNA binding-induced conformational change of the yeast transcriptional activator PRTE. *Cell* **62**, 367–377 (1990).
15. Cleary, M. A., Pendergrast, S. & Herr, W. Structural flexibility in transcription complex formation revealed by protein-DNA photocrosslinking. *Proc. Natl Acad. Sci. USA* **94**, 8450–8455 (1997).
16. Donaldson, L. W., Petersen, J. M., Graves, B. J. & McIntosh, L. P. Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J.* **15**, 125–134 (1996).
17. Werner, M. H. et al. The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation. *Cell* **38**, 761–771 (1995).
18. Jonsen, M. D., Petersen, J. M., Xu, Q. P. & Graves, B. J. Characterization of the cooperative function of inhibitory sequences in Ets-1. *Mol. Cell. Biol.* **16**, 2065–2073 (1996).
19. Skalkicky, J. J., Donaldson, L. W., Petersen, J. M., Graves, B. J. & McIntosh, L. P. Structural coupling of the inhibitory regions flanking the ETS domain of murine Ets-1. *Protein Sci.* **5**, 296–309 (1996).
20. Miner, J. N. & Yamamoto, K. R. Regulatory crosstalk at composite response elements. *Trends Biochem. Sci.* **16**, 426–426 (1991).
21. Wagner, S. & Green, M. R. DNA-binding domains: targets for viral and cellular regulatory. *Curr. Opin. Cell Biol.* **6**, 410–441 (1994).
22. Martin, M. L., Lieberman, P. M. & Curran, T. Fos-Jun dimerization promotes interaction of the basic region with TFIIE-34 and TFIIIE. *Mol. Cell. Biol.* **16**, 2110–2118 (1996).
23. Eckner, R., Yao, T., Oldread, E. & Livingston, D. M. Interaction and functional collaboration of p300/CBP and bHLH proteins in muscle and B-cell differentiation. *Genes Dev.* **10**, 2478–2490 (1996).
24. Sartorelli, V. et al. Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. *Mol. Cell. Biol.* **17**, 1010–1026 (1997).

25. Huang, C. C. & Herr, W. Differential control of transcription by homologous homeodomain coregulators. *Mol. Cell. Biol.* **16**, 2967–2976 (1996).
26. Misra, V. et al. Conformational alteration of Oct-1 upon DNA binding dictates selectivity in differential interactions with related transcriptional coactivators. *Mol. Cell. Biol.* **16**, 4404–4413 (1996).
27. Walker, S., Hayes, S. & O'Hare, P. Site-specific conformational alteration of the Oct-1 POU domain-DNA complex as the basis for differential recognition by Vmw65 (VP16). *Cell* **79**, 841–852 (1994).
28. Johansen, F. E. & Prywes, R. Two pathways for serum regulation of the c-fos serum response element require specific sequence elements and a minimal domain of serum response factor. *Mol. Cell. Biol.* **14**, 5920–5928 (1994).
29. Hill, C. S., Wynne, J. & Treisman, R. Serum-regulated transcription by serum response factor (SRF): a novel role for the DNA binding domain. *EMBO J.* **13**, 5421–5432 (1994).
30. Chen, J. D. & Evans, R. M. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* **377**, 454–457 (1995).
31. Mangelsdorf, D. J. et al. A direct repeat in the cellular retinol-binding protein type II gene confers differential regulation by RXR and RAR. *Cell* **66**, 555–561 (1991).
32. Kurokawa, R. et al. Polarity-specific activities of retinoic acid receptors determined by a co-repressor. *Nature* **377**, 451–454 (1995).
33. Kurokawa, R. et al. Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding. *Nature* **371**, 528–531 (1994).
34. Saaticioglu, F., Deng, T. & Karin, M. A novel cis element mediating ligand-independent activation by c-ErbA: implications for hormonal regulation. *Cell* **75**, 1095–1105 (1993).
35. Rahman, A., Esmaili, A. & Saaticioglu, F. A unique thyroid hormone response element in the human immunodeficiency virus type 1 long terminal repeat that overlaps the Sp1 binding sites. *J. Biol. Chem.* **270**, 31059–31064 (1995).
36. Tomie-Canie, M., Day, D., Samuela, H. H., Freedberg, I. M. & Blumenberg, M. Novel regulation of keratin gene expression by thyroid hormone and retinoid receptors. *J. Biol. Chem.* **271**, 1416–1423 (1996).
37. Härd, T. et al. Solution structure of the glucocorticoid receptor DNA-binding domain. *Science* **249**, 157–160 (1990).
38. Luisi, B. F. et al. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505 (1991).
39. Baumann, H. et al. Refined solution structure of the glucocorticoid receptor DNA-binding domain. *Biochemistry* **32**, 13463–13471 (1993).
40. van Tilborg, M. A. et al. Structure refinement of the glucocorticoid receptor-DNA binding domain from NMR data by relaxation matrix calculations. *J. Mol. Biol.* **247**, 689–700 (1995).
41. Schwabe, J. W. R. et al. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* **75**, 567–578 (1993).
42. Schwabe, J. W. et al. The estrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. *Structure* **3**, 201–213 (1995).
43. Dahlman-Wright, K. et al. Interaction of the glucocorticoid receptor DNA-binding domain with DNA as a dimer is mediated by a short segment of five amino acids. *J. Biol. Chem.* **266**, 3107–3112 (1991).
44. Drouin, J. et al. Homodimer formation is rate-limiting for high affinity DNA binding by glucocorticoid receptor. *Mol. Endocrinol.* **6**, 1299–1309 (1992).
45. Yamamoto, K. R., Pearce, D., Thomas, J. & Miner, J. N. In *Transcriptional Regulation* (eds McKnight, S. L. & Yamamoto, K. R.) 1169–1170 (Cold Spring Harbor Laboratory Press, NY, 1992).
46. Heck, S. et al. A distinct modulating domain in glucocorticoid receptor monomers in the repression of activity of the transcription factor AP-1. *EMBO J.* **13**, 4087–4095 (1994).
47. Lefstin, J. A., Thomas, J. R. & Yamamoto, K. R. Influence of a steroid receptor DNA-binding domain on transcriptional regulatory functions. *Genes Dev.* **8**, 2842–2856 (1994).
48. Cato, A. C. & Wade, E. Molecular mechanisms of anti-inflammatory action of glucocorticoids. *Bioassays* **18**, 371–378 (1996).
49. Drouin, J. et al. Novel glucocorticoid receptor complex with DNA element of the hormone-repressed POMC gene. *EMBO J.* **12**, 145–156 (1993).
50. König, H., Ponta, H., Rahmsdorf, H. J. & Herrlich, P. Interference between pathway-specific transcription factors: glucocorticoids antagonize phorbol ester-induced AP-1 activity without altering AP-1 site occupation in vivo. *EMBO J.* **11**, 2241–2246 (1992).
51. Diamond, M. I. et al. Transcription factor interactions: selectors of positive or negative regulation from a single DNA element. *Science* **249**, 1266–1271 (1990).
52. Jonat, C. et al. Antitumor promotion and antiinflammation: down-modulation of AP-1 (fos/jun) activity by glucocorticoid hormone. *Cell* **62**, 1189–1204 (1990).
53. Yang-ye, H.-F. et al. Transcriptional interference between c-Jun and the glucocorticoid receptor: mutual inhibition of DNA binding due to direct protein-protein interaction. *Cell* **62**, 1205–1215 (1990).
54. Starr, D. B. et al. Intracellular receptors use a common mechanism to interpret signaling information at response elements. *Genes Dev.* **10**, 1271–1283 (1996).
55. Sauddek, V. et al. Solution structure of the basic region from the transcriptional activator GCN4. *Biochemistry* **30**, 1310–1317 (1991).
56. Ellenberger, T. E. et al. The GCN4 basic leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* **71**, 1223–1237 (1992).
57. O'Neil, K. T., Shuman, J. D., Ampe, C. & DeGrado, W. F. DNA-induced increase in the alpha-helical content of C/EBP and GCN4. *Biochemistry* **30**, 9030–9034 (1991).
58. Krebs, D. et al. The basic subdomain of the c-Jun oncoprotein. A joint CD, Fourier-transform infrared and NMR study. *Eur. J. Biochem.* **231**, 370–380 (1995).
59. Ebnet, A. et al. Biophysical characterization of the c-Myb DNA-binding domain. *Biochemistry* **33**, 14586–14593 (1994).
60. Ogata, K. et al. The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nature Struct. Biol.* **3**, 178–187 (1996).
61. Jamin, N. et al. Secondary structure of the DNA-binding domain of the c-Myb oncoprotein in solution. A multidimensional double and triple heteronuclear NMR study. *Eur. J. Biochem.* **216**, 147–154 (1993).
62. Carr, M. D. et al. Structure of the B-Myb DNA-binding domain in solution and evidence for multiple conformations in the region of repeat-2 involved in DNA binding: implications for sequence-specific DNA binding by Myb proteins. *Eur. J. Biochem.* **235**, 721–735 (1996).
63. Fujita, T., Nolan, G. P., Ghosh, S. & Baltimore, D. Independent modes of transcriptional activation by the p50 and p65 subunits of NF-kappa B. *Genes Dev.* **6**, 775–787 (1992).
64. Hay, R. T. & Nicholson, J. DNA binding alters the protease susceptibility of the p50 subunit of NF-kappaB. *Nucleic Acids Res.* **21**, 4592–4598 (1993).
65. Matthews, J. R. et al. Conformational changes induced by DNA binding of NF-kappa B. *Nucleic Acids Res.* **23**, 3393–3402 (1995).
66. Toney, J. H. et al. Conformational changes in chicken thyroid hormone receptor alpha 1 induced by binding to ligand or to DNA. *Biochemistry* **32**, 2–6 (1993).

Acknowledgements. We thank W. Herr, C. Jamieson, A. Johnson, E. Levine, D. B. Starr and J. Weissman for comments on the manuscript. The authors are supported by grants from the NIH and the NSF.

The paradox of drowned carbonate platforms and the origin of Cretaceous Pacific guyots

Paul A. Wilson*, Hugh C. Jenkyns†, Henry Elderfield* & Roger L. Larson‡

* Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

† Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK

‡ Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882-1197, USA

Geochemical, stratigraphic and palaeolatitudinal data from deep boreholes drilled through Pacific guyots—flat-topped seamounts—help to explain the drowning of these Cretaceous shallow-water carbonate platforms that once thrived through the accumulation of biogenic and inorganic calcium carbonate sediment in mid-oceanic regions. The platforms drowned sequentially over a 60-million-year interval while they were being transported northward by Pacific plate motion through a narrow equatorial zone (~0–10° S). Such platforms were apparently resistant to the effects of Cretaceous oceanic anoxic events. Although the mechanism responsible for drowning remains unknown, the tropics have not always been the refuge for atolls that they are today.

Oceanic volcanic islands subside as they age to produce chains of coral atolls¹ and seamounts² that provide important clues to Earth's tectonic history³. The north Pacific Ocean is strewn with flat-topped seamounts or 'guyots'² of Cretaceous age (~150–65 Myr ago), an inferred 'greenhouse' period of exceptional global oceanic volcanism, high concentrations of atmospheric CO₂, climate warmth, comparative sea-level elevation and shelf-carbonate formation^{4–8}. Originally, guyots were assumed to be erosionally truncated drowned volcanoes². However, subsequent dredging has shown that many guyots are drowned shallow-water carbonate platforms^{4,5}. Drowned platforms of this type are common features in the geological record and thus present a fundamental paradox because the growth potential of healthy platforms of this type is thought to be at least as great—possibly one or two orders of magnitude greater than—long-term rates of relative sea-level rise⁹. Initial attempts⁹ to resolve this paradox invoked short-term events

of (1) anomalously rapid rise in relative sea level (10 m kyr⁻¹; kyr timescale) and/or (2) environmental-stress-induced deterioration in the potential for platform growth. More recent studies have addressed the well-known inverse relationship between sedimentation rates and the timespan over which these rates are measured, and have suggested that slower rates of relative sea-level rise acting persistently over longer periods (0.1 m kyr⁻¹; Myr timescale) may also favour platform drowning¹⁰. In the case of the Cretaceous, the palaeoceanographic effects of the well-known periods of oceanic anoxia¹¹ have been widely invoked to explain carbonate-platform drowning^{12–15}.

Ocean Drilling Program (ODP) Legs 143 and 144 drilled a series of holes into a total of seven guyots spanning approximately 30° of latitude and 35° degrees of longitude in the western Pacific Ocean (Fig. 1). The sequences recovered from six of these guyots have three components: an underlying volcanic edifice, a shallow-water car-

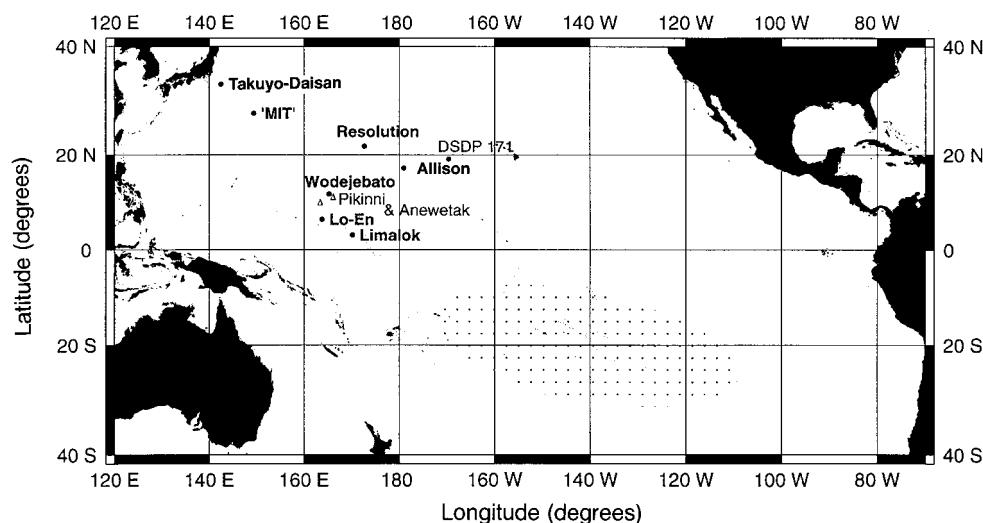


Figure 1 Map of the Pacific Ocean showing the present location of the seven guyots (bold type) drilled during ODP Legs 143 and 144. Also shown are the atolls (triangles) of Anewetak (formerly Enewetak) and Pikinni (formerly Bikini) and Deep Sea Drilling Program (DSDP) Site 171. The volcanic edifices of all of these guyots are thought to have formed within the South Pacific isotopic and thermal

anomaly (stippled) region (see text). All of the guyots shown have well-developed drowned shallow-water carbonate platforms except Lo-En. Five of these six drowned carbonate platforms are Cretaceous in age whereas Limalok is Palaeogene (see text).

bonate platform and an overlying 'cap' of pelagic sediments (the remaining guyot, Lo-En, is composed of a pelagic cap and a volcanic edifice but no true carbonate platform). The six shallow-water carbonate platforms drilled contain abundant rudists (cemented bivalves, possibly reef-forming) and/or corals, calcareous algae, benthonic foraminifera and ooids (sand-sized grains composed of a nucleus and a precipitated carbonate coating) and can be thought of as isolated drowned ancient analogues of the modern Bahama banks, albeit situated on oceanic crust. Two end-member competing hypotheses have emerged to explain the paradox of carbonate-platform drowning as manifested by the guyots drilled during ODP Legs 143 and 144. The first hypothesis, 'death-by-emergence-and-submergence', suggests that carbonate-platform drowning was related to regional tectonic or eustatic effects that allowed a short-term pulse in relative sea-level rise to outpace carbonate-platform production^{16–18}. Specifically, this mechanism invokes a period of low sea level, exposure of the carbonate platform, karstification (dissolution of the limestone surface by meteoric water) and a subsequent rapid rise in relative sea level. In this model, carbonate-platform growth during rapid transgression is thought to have been hampered by antecedent karstic topography, possibly because high rates of sediment production are required from areally restricted topographic highs to fill the additional accommodation space created by karstic topographic lows. The second hypothesis, 'death-in-the-tropics', suggests that drowning was related to plate-tectonic transport of platforms through low-latitude waters that were in some way harmful to the accumulation of shallow-water carbonate sediments¹⁹. An additional problem confronting both of the above hypotheses is that of the close proximity of surviving atolls (for example, Anewetak and Bikini) to guyots (for example,

Limalok, Lo-En and Wodejebato) in the Marshall Islands, suggesting that juxtaposed shallow-water carbonate platforms can apparently undergo radically different growth/drowning histories (Fig. 1).

Here we present geochemical, stratigraphic and palaeolatitudinal data from the carbonate platforms drilled during ODP Legs 143 and 144. Based on the stable-isotope geochemistry of calcite cements from the upper 150 m of the drilled platforms, we find no evidence to support the hypothesis that these platforms were subjected to a period of intense karstification before drowning. Chemostratigraphic results indicate that these platforms drowned sequentially over a long interval of Cretaceous and Palaeogene time (~111–48 Myr ago). Furthermore, application of these drowning ages to a palaeomagnetic reconstruction of Pacific plate motion suggests that all of the drilled platforms drowned in a narrow zone (~0–10° S) asymmetrically about the Equator. These findings imply that surface waters in this low-latitude zone were, at times, in some way harmful to carbonate-platform sedimentation. Our data show that, contrary to popular belief, mid-oceanic carbonate platforms were resistant to the palaeoceanographic effects of Cretaceous oceanic anoxic events. Nevertheless, the mechanism responsible for drowning remains poorly constrained. Regardless of the palaeoceanographic mechanism responsible, however, our results indicate that, at certain times in the past, mid-oceanic carbonate platforms were jeopardized by plate-tectonic transport through low-latitude waters, an apparent reversal of fortunes in comparison to modern atolls that thrive in low-latitude waters but are threatened by transportation to high latitudes^{1,3}. This finding has important implications for the continuing debate over the proposed palaeoceanographic controls on changes in the biological diversity and extinction rates of tropical reefs throughout greenhouse periods^{20,21}.

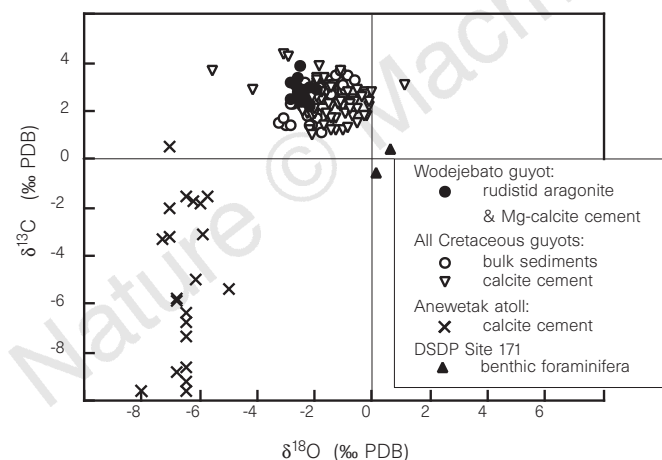


Figure 2 Stable carbon- and oxygen-isotope data. All data are referenced to the Pee Dee belemnite (PDB) carbonate standard, and are from the upper 150 m of the Cretaceous shallow-water carbonate platforms on: Resolution, Allison, Wodejebato, 'MIT' and Takuyo-Daisan Guyots (Fig. 1). Data^{40,41} are also shown for (1) meteoric-water calcite cements from the Pleistocene sequences (upper ~80 m) of Anewetak atoll and (2) deep-sea benthic foraminiferal calcite from the lower Maastrichtian sequence at DSDP Site 171 (Fig. 1). The guyot data indicate, first, that diagenesis of the top of these platforms was dominated by marine rather than meteoric waters and, second, that there is no geochemical evidence to support the hypothesis that these platforms were subjected to significant meteoric-water diagenesis before drowning (see text). This finding is consistent with a number of independent studies of the isotope geochemistry of cements from individual ODP 143 and 144 guyots. Methods are those of refs 31, 36, 37. Wodejebato measurements were made using a Finnigan Mat 251 mass spectrometer with a Keil Carbonate device (precision ~δ¹³C = 0.01‰, δ¹⁸O = 0.04‰, PDB). Measurements from the other guyots were made using a VG Prism mass spectrometer (precision for both carbon- and oxygen-isotope ratios was better than 0.1‰).

Origin of Pacific guyots

In principle, the two end-member hypotheses for the origin of the ODP Legs 143 and 144 guyots are relatively straightforward to test because they make fundamentally different predictions regarding carbonate-platform drowning: (1) 'death-by-emergence-and-submergence' predicts simultaneous platform karstification and drowning events across a wide range of latitudes; (2) 'death-in-the-tropics' predicts sequential drowning of non-karstified platforms at low latitude. Therefore, in order to test these two hypotheses, the timing and location of platform drowning and the mode of platform-top diagenesis must be established. 'Death-by-emergence-and-submergence' has been proposed^{16–18} on the basis of the following observations. First, rough, supposedly karstic carbonate-platform topographies, second, petrographical evidence for dissolution by meteoric water (for example, mouldic and vuggy pores (cavities produced by the removal of readily soluble fossils and matrix)), and third, biostratigraphically inferred simultaneous ages for drowning. In contrast, 'death-in-the-tropics' has been proposed¹⁹ on the basis of biostratigraphically inferred sequential ages for drowning. However, these lines of evidence are equivocal for two reasons. (1) Rough topographies and mouldic porosity can be produced on the tops of carbonate platforms by processes other than karstification (for example, physical erosion²² and aragonite dissolution in even relatively shallow, ~350 m, submarine waters^{23–25}); and (2) biostratigraphic methods, relying primarily on environmentally dependent benthic foraminifera for dating shallow-water carbonate sediments, are associated with considerable uncertainties^{23,26}. For these reasons, we use strontium-isotope (⁸⁷Sr/⁸⁶Sr) stratigraphy to date shallow-water carbonate sediments of the guyots, and the stable-isotope (δ¹³C and δ¹⁸O) geochemistry of cements to constrain the mode of diagenesis in platform-top sediments. Previous studies of Cenozoic atolls have shown that such techniques are ideally suited to mid-oceanic carbonate-platform facies^{23,27–29} and application of the Sr-isotope technique to Cretaceous

sections^{30–32} has recently become possible through the advent of reliable seawater reference curves for this period^{33–35}. One important reason why Sr-isotope stratigraphy is ideally suited to mid-oceanic carbonate platform sections is that such platforms are virtually isolated from continentally derived radiogenic sources of strontium. Such isolation usually ensures that primary metastable carbonate minerals (aragonite and magnesian calcite) are diagenetically altered to secondary calcite in pore fluids having an identical Sr-isotope composition to that of sea water at the time of deposition^{23,27–29}. Extensive work on bulk sediments and co-existing microsampled separates of depositional and diagenetic carbonate (including samples of aragonite, magnesian calcite and low-magnesian calcite) has shown that, except in rare circumstances, relatively small (~2 mg), bulk-rock samples from Cretaceous guyots, carefully selected for the homogeneous extent of lithification, retain their Sr-isotopic and hence chemostratigraphic integrity^{30–32,36,37}.

Calcite-cement geochemistry and platform karstification

To test the hypothesis that the guyot platforms were subjected to significant karstification before drowning, we present (in Fig. 2) $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ data for bulk limestones (open circles) and microsampled separates from the upper 150 m of all five of the Cretaceous carbonate platforms (~200–1,600 m thick) drilled: Allison, Resolution, Wodejebato, 'MIT' and Takuyo-Daisan. Data from microsampled material fall into two groups: (1) Magnesian-calcite cements and encased rudistid aragonite (solid circles in Fig. 2)

from a thin (~40 cm) bed of rudstones from Wodejebato and (2) low-magnesian calcite (hereafter, 'calcite') cements (open triangles in Fig. 2) from skeletal wackestones from all of the guyots. The exceptional textural, mineralogical and geochemical preservation of the group 1 samples indicates that the magnesian-calcite cements precipitated within Wodejebato from shallow-marine waters^{36,37}. Stable-isotope data from these magnesian-calcite cements and encased aragonitic rudists fall in a narrow range (Fig. 2) and indicate that these metastable minerals, like analogues from the living reefs of modern atolls, formed in near-stable isotopic equilibrium with their parent shallow-marine waters (Late Cretaceous, $\sim 69 \pm 1$ Myr)³⁶. Therefore, this narrow range of stable-isotope values (solid circles, Fig. 2) provides a syndepositional point of reference from which to constrain the post-depositional diagenetic history of all of the Cretaceous guyots and thus evaluate the karstification hypothesis.

The calcite cements from the limestones of all five Cretaceous guyots (group 2 samples) have $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values in a similar range to the bulk-carbonate samples (Fig. 2). Furthermore, virtually all of these cements and bulk carbonates have marginally higher $\delta^{18}\text{O}$ and lower $\delta^{13}\text{C}$ values than the magnesian calcites from Wodejebato (Fig. 2). These data indicate that diagenesis within all five of these Cretaceous carbonate platforms was dominated by fluids that had higher $\delta^{18}\text{O}$ values and/or lower temperatures (and lower $\delta^{13}\text{C}$ values) than surface waters of the Late Cretaceous western Pacific Ocean³⁷. Such fluids must have been marine rather

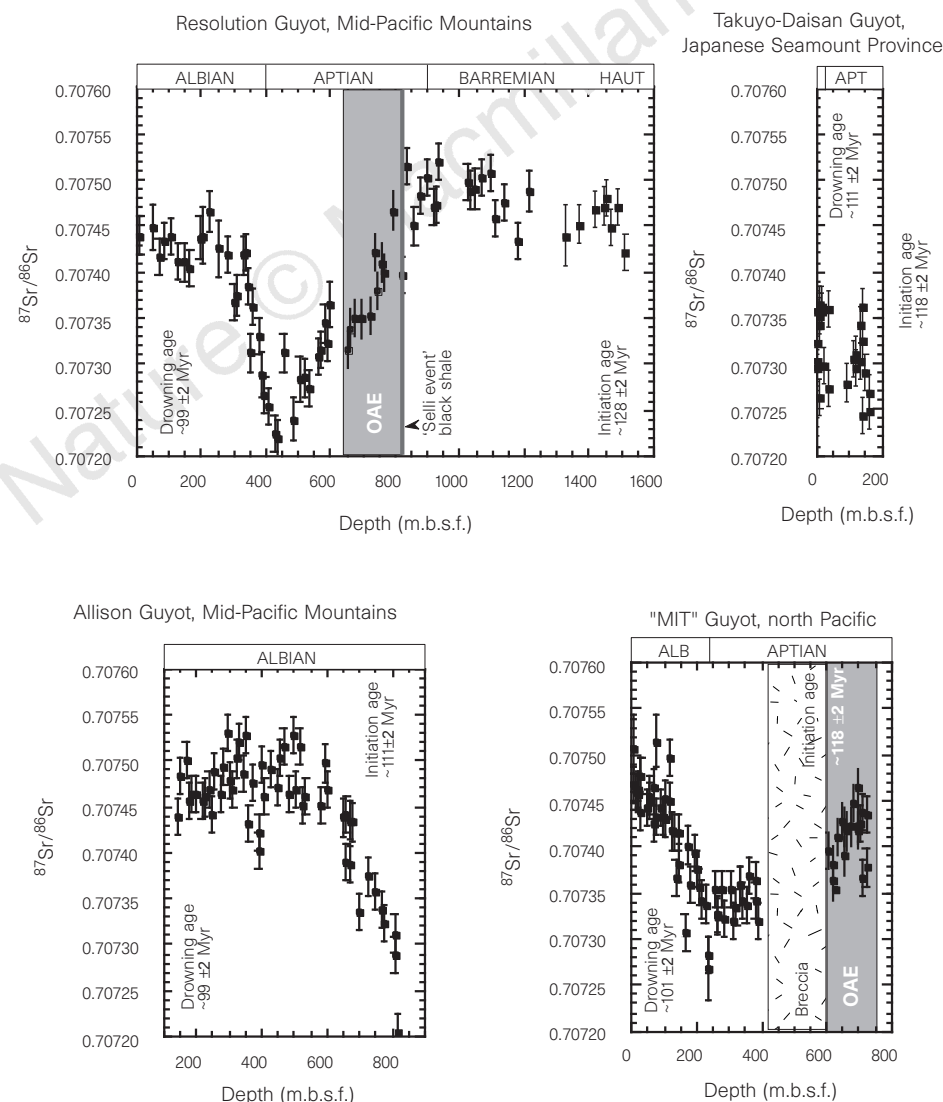


Figure 3 Sr-isotope ($^{87}\text{Sr}/^{86}\text{Sr}$) stratigraphies for the Cretaceous carbonate platforms of Resolution, Allison, 'MIT' and Takuyo-Daisan guyots. Stratigraphy for Resolution guyot is modified from original interpretation³⁰. This platform contains evidence (the 'Selli-event' black shale⁵⁰) that mid-oceanic carbonate platforms such as these were resistant to the palaeoceanographic effects of oceanic anoxic events (OAEs). The associated well-known Aptian double positive $\delta^{13}\text{C}$ excursions are recorded in the carbonate sediments from both Resolution and 'MIT' guyots. Sr-isotope data from the volcanoclastic breccia (~200 m thick) within the sequence at 'MIT' have been omitted, Sr-isotope data in this study were generated in three different laboratories, at the Universities of Cambridge, Oxford and North Carolina. To ensure comparability, all Sr-isotope data were normalized to a value of NBS 987 = 0.710250 (value determined at North Carolina over a period during which analyses of Allison and Resolution guyots were undertaken = 0.710250 ± 22 , $n = 104$; value determined at Oxford over the period during which analyses of Resolution guyot were undertaken = 0.710260 ± 20 , $n = 27$; values determined at Cambridge during the two periods during which analyses of Allison, Resolution, 'MIT' and Takuyo-Daisan guyots were undertaken = 0.710251 ± 21 , $n = 139$ and 0.710253 ± 16 , $n = 76$). Error bars shown on individual data points are as $\pm 20 \times 10^{-6}$ for all analyses unless precision was worse than this, in which case the greater uncertainty is shown.

than meteoric. Despite the low latitudes of modern Pacific atolls such as Anewetak and Bikini (Fig. 1), local meteoric water has significantly lower $\delta^{18}\text{O}$ values (~ -5.8 to -3.8‰ , SMOW) than local surface sea water (~ 0.0 to $+0.5\text{‰}$, SMOW)^{36,38} because of air-mass convection to extreme altitudes associated with the Inter Tropical Convergence Zone³⁹. Furthermore, calcite cements precipitated from meteoric waters in the upper 80 m of Anewetak during Pleistocene exposure of this atoll have much lower stable-isotope values⁴⁰ than any of the calcite cements analysed from the upper 150 m of the five Cretaceous guyots (Fig. 2). Therefore, we find no geochemical evidence to support the hypothesis that the guyot carbonate platforms were subjected to intense dissolution and lithification in meteoric waters, 'karstification', before drowning. In fact, comparison with stable-isotope data⁴¹ from deep-sea benthic foraminiferal calcite in the lower Maastrichtian sequence at DSDP Site 171 suggests that the calcite cements from the Cretaceous guyots precipitated from sea water within or beneath the oceanic thermocline (Fig. 2).

Timing and location of platform drowning

In Fig. 3 we show Sr-isotope stratigraphies of the drowned Cretaceous carbonate platforms for Resolution, Allison, 'MIT' and Takuyo-Daisan guyots. Stage boundaries and ages for the lowermost and uppermost shallow-water platform sediments (estimated dates for the timing of initiation and cessation of shallow-water carbonate-platform deposition) shown in Fig. 3 are assigned using the appropriate Sr-isotope seawater reference curves^{33–35} and timescale⁴². In general, the Sr-isotope dates are consistent with biostratigraphic interpretations where well-preserved planktonic microfossils and nannofossils are present²⁶. However, significant discrepancies exist between the Sr-isotope and benthic biostratigraphic estimates for the timing of cessation of shallow-water

carbonate platform deposition. For example, Sr-isotope stratigraphy indicates that Takuyo-Daisan drowned (early Albian, ~ 111 Myr ago) before 'MIT' (early late Albian, ~ 101 Myr ago) whereas benthic foraminiferal stratigraphy^{26,43} suggests that 'MIT' (late Albian age) drowned before Takuyo-Daisan (latest Albian age). However, benthic foraminifera are known to be facies-dependent and hence of uncertain stratigraphic value outside of their type area (for example, southern France: Urgonian facies⁴³). Therefore our best estimates for the age of platform drowning are provided by the Sr-isotope dates for the cessation of shallow-water carbonate platform sedimentation. Sr-isotope stratigraphies for Limalok and Wodejebato are published elsewhere and the estimated ages for platform initiation and drowning are respectively: Limalok $\sim 57.5 \pm 2.5$, $\sim 48 \pm 2$ Myr ago; Wodejebato $\sim 76 \pm 1$, $\sim 69 \pm 1$ Myr ago (refs 26, 31, 32).

Palaeomagnetic and geochemical data⁴⁴ on basaltic rocks from ODP Leg 143 and 144 boreholes indicate that the volcanic edifices of all of the drilled guyots formed within the South Pacific isotopic and thermal anomaly (stippled region, Fig. 1). Carbonate reefs probably fringed these edifices as soon as they erupted into the photic zone¹. However, our only stratigraphic records come from the drill sites, generally located towards the platform interiors. These records show that, in some cases (for example, Limalok, Wodejebato and Takuyo-Daisan), the main phase of shallow-water carbonate sedimentation did not begin on the guyots until some time after basalt emplacement because of the significant periods of volcanic-island emergences (~ 3 – 7 Myr)¹⁹. In Table 1 we summarize our age estimates for platform initiation and platform drowning, apply these dates to a palaeomagnetic reconstruction of Pacific plate motion (based on seamount magnetization for the Late Cretaceous⁴⁵ and magnetic anomaly skewness for the Early Cretaceous⁴⁶) and thereby also estimate the palaeolatitudes of platform initiation and drowning. Figure 4 shows that the demise of these Pacific guyot platforms cannot be attributed to a single event because the shallow-water carbonate platforms drowned sequentially across a wide range of Cretaceous and Palaeogene time (~ 60 million years). Furthermore, although the main phase of shallow-water carbonate sedimentation on each guyot began well within the tropics of the southern hemisphere (Table 1), our results show that all six platforms (Allison, Resolution, Limalok, Wodejebato,

Table 1 Guyots studied

Guyot	Age (Myr)	Palaeolatitude (deg.)
Limalok		
Initiation	$\sim 57.5 \pm 2.5$	10.80 ± 2.50 S
Drowning	$\sim 48.0 \pm 2$	8.15 ± 2.75 S
Present		5.55 N
Wodejebato		
Initiation	$\sim 76 \pm 1$	11.60 ± 3.20 S
Drowning	$\sim 69 \pm 1$	6.20 ± 2.20 S
Present		11.90 N
'MIT'		
Initiation	$\sim 119 \pm 2$	11.55 ± 3.45 S
Drowning	$\sim 101 \pm 2$	8.35 ± 3.35 S
Present		27.30 N
Takuyo-Daisan		
Initiation	$\sim 118 \pm 2$	4.80 ± 3.50 S
Drowning	$\sim 111 \pm 2$	3.40 ± 3.50 S
Present		34.15 N
Allison		
Initiation	$\sim 111 \pm 2$	13.00 ± 3.30 S
Drowning	$\sim 99 \pm 2$	11.75 ± 3.25 S
Present		18.45 N
Resolution		
Initiation	$\sim 128 \pm 2$	14.50 ± 3.30 S
Drowning	$\sim 99 \pm 2$	10.70 ± 3.30 S
Present		21.35 N

Summarized age and palaeolatitude estimates for the main phase of platform initiation (see text) and platform drowning for all six guyots having well-developed shallow-water sedimentary sequences. Present latitudes of these guyots are also given for comparison. Age estimates for 'MIT', Takuyo-Daisan, Allison and Resolution are from Fig. 3. Age estimates for Limalok and Wodejebato are published elsewhere^{26,31,32}. The estimated palaeolatitudes for drowning are calculated using the same basic method as in ref. 19 (using a model for plate motion based on seamount magnetization for the Late Cretaceous⁴⁵ and magnetic anomaly skewness for the Early Cretaceous⁴⁶). However, we now use the M5 to M10 pole⁴⁵ (~ 128 – 132 Myr) instead of the M0 to M4 pole solution for the old end member because the new pole has a smaller error ellipse than for the previous one and is consistent with an independently computed⁴² pole for 129 Myr. The skewness pole assumes no anomalous skewness, but also has an error ellipse that overlaps most of the error ellipse for the same-aged pole calculated with anomalous skewness¹⁹. Thus, use of the latter pole would only change the older palaeolatitudes by $\sim 1^\circ$ to the south.

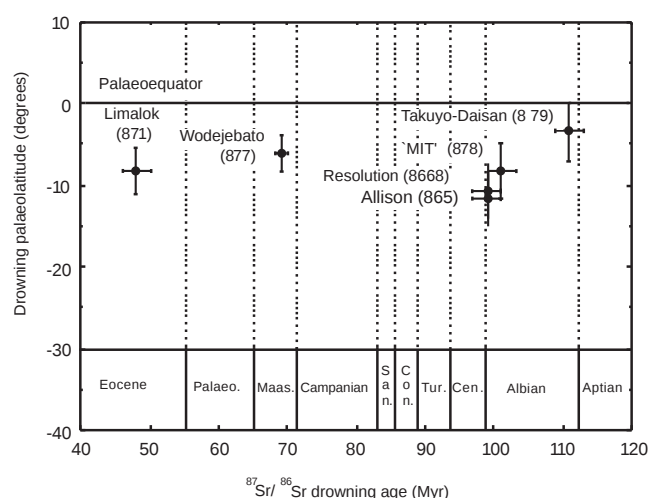


Figure 4 The timing and location of platform drowning. Data were estimated from the Sr-isotope ages for the cessation of carbonate-platform deposition on the guyots and the history of Pacific plate motion (see Fig. 3 and Table 1). Results support the hypothesis, 'death-in-the-tropics', that the platforms drowned as a result of the tectonic transport of platforms through low-latitude waters that were in some way harmful to the accumulation of shallow-water carbonate sediments (see text).

'MIT' and Takuyo-Daisan) drowned as they approached the palaeoequator at latitudes in which modern atolls continue to thrive ($\sim 0\text{--}10^\circ\text{S}$, Fig. 4). These findings, together with the lack of geochemical evidence for meteoric-water diagenesis of these platforms, are fundamentally incompatible with the predictions of 'death-by-emergence-and-submergence' but are consistent with the predictions of 'death-in-the-tropics': namely, platform drowning was related to the plate-tectonic transport of platforms through low-latitude waters that were in some way harmful to shallow-water carbonate accumulation.

Surviving atolls and platform erosion

The above findings may also explain the existence of surviving atolls in close proximity to guyots in the Marshall Islands (Fig. 1). Anewetak is the only one of these atolls that has been drilled to volcanic basement and can therefore be evaluated. The top of the volcanic edifice on Anewetak is of Late Cretaceous age ($\sim 75 \pm 0.6$ Myr ago)⁴⁷ whereas the basal overlying shallow-water carbonate platform is apparently much younger (middle to late Eocene, $\sim 40 \pm 5$ Myr ago)^{23,47}. Therefore, while Limalok was drowning at $\sim 8^\circ\text{S}$ latitude during the early mid-Eocene (~ 48 Myr ago, Fig. 4), nearby Anewetak was merely a volcanic island and thus immune to passage through the palaeolatitudinal zone ($\sim 0\text{--}10^\circ\text{S}$) of platform drowning. Furthermore, our calculations indicate that by the time that Anewetak began to accumulate shallow-water carbonate sediments, it was located at the Equator. Since this time, Anewetak has continued to thrive as an atoll during the northward transition to its present latitude ($\sim 11.5^\circ\text{N}$, Fig. 1). This observation implies that the low-palaeolatitudinal zone of platform drowning perhaps did not extend into the Northern Hemisphere.

The possibility that the tops of the guyot shallow-water carbonate platforms may have been subjected to significant erosion cannot be discounted. Therefore, calculated ages and latitudes for platform drowning must be regarded as maximum estimates (oldest dates; highest latitudes). Nevertheless, five lines of evidence suggest that erosion-induced uncertainties in estimation of these ages and latitudes are relatively small and that our interpretation of these data is therefore robust: (1) The latitudinal range of the guyots today ($\sim 30^\circ$, Fig. 1) is large in comparison to our estimated latitudinal range for platform drowning ($\sim 10^\circ$, Fig. 4). (2) Estimated long-term rates of lateral plate motion are low (maximum $\sim 1.0^\circ$ latitude Myr^{-1}) in comparison to estimated long-term average sedimentation rates (maximum ~ 150 m Myr^{-1} for the drowned shallow-water carbonate platforms, thereby requiring the removal of substantial section thicknesses in order to prejudice our interpretations). (3) The tops of the guyots lie in water depths ($\sim 1,200\text{--}1,500$ m) significantly shallower than those ($\sim 3,000$ m) at which calcite dissolution can contribute to platform erosion (in fact, these carbonate platforms are typically capped by well-preserved pelagic calcite sediments). (4) The Sr-isotope date for the cessation of platform growth at 'MIT' cannot be distinguished from the date of the onset of pelagic sedimentation on this guyot²⁶. (5) Extreme heterogeneity of lithification in shallow-water carbonate sediments favours erosional dissection over planation and, thus, the topographic retention of primary depositional surfaces⁴⁸.

Drowning mechanisms

Our data indicate that the guyot carbonate platforms drowned sequentially across a wide range of time while they were being transported northwards by Pacific plate motion through a narrow palaeolatitudinal zone ($\sim 0\text{--}10^\circ\text{S}$, Fig. 4). These findings imply that surface waters in this palaeolatitudinal zone were, in some way, harmful to shallow-water carbonate sedimentation. Extremes of water temperature and nutrient levels are perhaps the most frequently invoked parameters for harmful water quality and demise of carbonate platforms⁹. Elsewhere, increases in nutrient levels and

decreases in sea-surface temperature associated with acceleration of the 'biological pump' and atmospheric CO_2 draw-down during oceanic anoxic events have been widely invoked to explain Cretaceous carbonate-platform drowning^{12–15}. Furthermore, the palaeolatitudinal zone of platform drowning ($\sim 0\text{--}10^\circ\text{S}$) corresponds remarkably well with the equatorially asymmetric zone of maximum low-latitude upwelling intensity and, thus, surface-water nutrient levels in the modern Pacific Ocean (maximum nutrient levels $\sim 0\text{--}8^\circ\text{S}$)⁴⁹. However, our data show that none of the guyot platforms examined here drowned during a major Cretaceous oceanic anoxic event. In fact, it seems that such platforms were resistant to the palaeoceanographic effects of oceanic anoxic events because the carbon-isotope stratigraphies of both Resolution and 'MIT' record the 'double pronged' Aptian positive $\delta^{13}\text{C}$ excursion associated with oceanic anoxic event 1 (OAE-1) and the platform at Resolution even contains the correlative 'Selli event' black shale⁵⁰ (Fig. 3).

The above observations imply that high nutrient levels (assumed to promote OAEs¹¹) and/or associated periods of relatively low sea-surface temperature were unlikely to have been responsible for the inferred harmful nature of the low-latitude waters in which the platforms drowned. However, the duration of the Selli event was probably relatively short (between ~ 1.0 and 0.5 Myr). Therefore, despite the fact that Resolution and 'MIT' experienced relatively rapid average rates of subsidence during Aptian time (maximum rates: Resolution ~ 60 m Myr^{-1} ; 'MIT' ~ 40 m Myr^{-1} , Fig. 3), the apparent resistance of these two platforms to OAE-1 might be explained by post-Selli-event platform recovery from within the photic zone ('MIT' $\sim 20\text{--}40$ m below sea level, m.b.s.l.; Resolution $\sim 30\text{--}60$ m.b.s.l.). Thus, we cannot rule out the possibility that the drowning of the Leg 143 and 144 platforms was caused by longer-term exposure of platforms to less extreme levels of nutrient excess associated with Cretaceous equatorial productivity. Similarly, by analogy with modern coral-reef systems⁵¹, we do not discount the possibility that exposure to abnormally high sea-surface temperatures was responsible for drowning these platforms. In fact, basic theories of tropical atmospheric and oceanic dynamics predict that low-latitude sea surface temperatures were warmer (by $\sim 2\text{--}5^\circ\text{C}$) than today during the Cretaceous 'greenhouse'^{6–8}, and $\delta^{18}\text{O}$ -palaeothermometry³⁶ of the well-preserved magnesian calcites and encased rudistid aragonite from Wodejebato suggests that this was the case for the western equatorial Pacific during the Late Cretaceous (~ 69 Myr ago). However, considerable uncertainties are associated with the use of modern analogues to infer the environmental tolerances of ancient (in the case of rudists, extinct) organisms. In fact, it has been hypothesized that episodic production of extremely warm ($>30^\circ\text{C}$) low-latitude surface waters favoured the proliferation of rudistid bivalves and the expansion of tropical carbonate platforms during the Cretaceous²⁰. Although this specific hypothesis is the subject of current debate²¹, parameters such as low-latitude sea surface temperature and nutrient concentration undoubtedly must have varied in both time and space during the Cretaceous. Ultimately, therefore, individual drowning events were a local function of the interplay among water quality, potential platform growth and subsidence (Table 1, Figs 3 and 4). For the above reasons, the mechanism responsible for the drowning of the carbonate platforms drilled during ODP Legs 143 and 144 remains enigmatic. Nevertheless, our results imply that the survival of Cretaceous and Palaeogene mid-oceanic shallow-water carbonate platforms was threatened by, rather than promoted by, tectonic passage through low-latitude waters. □

Received 22 September 1997; accepted 9 March 1998.

1. Darwin, C. *The Structure and Distribution of Coral Reefs* (Smith, Eder & Co., London, 1842).
2. Hess, H. H. Drowned ancient islands of the Pacific basin. *Am. J. Sci.* **244**, 772–791 (1946).
3. Wilson, J. T. Evidence from islands on the spreading of ocean floors. *Nature* **197**, 536–538 (1963).
4. Hamilton, E. L. Sunken islands of the Mid-Pacific mountains. *Mem. Geol. Soc. Am.* **64**, (1956).
5. Menard, H. W. & Ladd, H. S. Oceanic islands, seamounts, guyots and atolls. *The Sea* **3**, 365–385 (1963).

6. Schlanger, S. O., Jenkyns, H. C. & Premoli Silva, I. Volcanism and vertical tectonics in the Pacific basin related to the global Cretaceous transgressions. *Earth Planet. Sci. Lett.* **52**, 435–449 (1981).
7. Barron, E. J. A warm, equable Cretaceous: the nature of the problem. *Earth Sci. Rev.* **19**, 305–338 (1983).
8. Larson, R. L. Latest pulse of Earth: Evidence for a mid-Cretaceous superplume. *Geology* **19**, 547–550 (1991).
9. Schlager, W. The paradox of drowned reefs and carbonate platforms. *Bull. Geol. Soc. Am.* **92**, 197–211 (1981).
10. Schlager, W., Marsal, D., van der Geest, P. A. G., Sprenger, A. & Fouke, B. W. Platform drowning, scaling of sedimentation rates and environmental change. *Soc. Econ. Paleont. Mineral. Congr. Sed. Geol.* **1**, 110 (1995).
11. Arthur, M. A., Jenkyns, H. C., Brumsack, H.-J. & Schlanger, S. O. in *Cretaceous Resources, Events and Rhythms* (eds Ginsburg, R. N. & Beaudoin, B.) 75–119 (NATO ISI Ser. 304, Kluwer, Dordrecht, 1990).
12. Hallock, P. & Schlager, W. Nutrient excess and the demise of coral reefs and carbonate platforms. *Palaios* **1**, 389–398 (1986).
13. Vogt, P. R. Volcanogenic upwelling of anoxic, nutrient-rich water: a possible factor in carbonate-bank/reef demise and benthic faunal extinctions? *Bull. Geol. Soc. Am.* **101**, 1225–1245 (1989).
14. Philip, J. M. & Airaud-Crumière, C. The demise of the rudist-bearing carbonate platforms at the Cenomanian/Turonian boundary: a global control. *Coral Reefs* **10**, 115–125 (1991).
15. Rougerie, F. & Fagerstrom, J. A. Cretaceous history of Pacific Basin guyots reefs: a reappraisal based on geothermal endo-upwelling. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **112**, 239–260 (1994).
16. Winterer, E. L., van Waasbergen, R., Mammerickx, J. & Stuart, S. Karst morphology and diagenesis of the top of Albian limestone platforms, mid-Pacific Mountains. *Proc. ODP B Sci. Res.* **143**, 433–470 (1993).
17. Winterer, E. L. & Sager, W. W. Synthesis of drilling results from the mid-Pacific Mountains: Regional context and implications. *Proc. ODP B Sci. Res.* **143**, 497–535 (1993).
18. Winterer, E. L. & Metzler, C. V. Origin and subsidence of guyots in Mid-Pacific mountains. *J. Geophys. Res.* **89**, 9969–9979 (1984).
19. Larson, R. L., Erba, E., Nakanishi, M., Bergersen, D. D. & Lincoln, J. M. Stratigraphic, vertical subsidence and paleolatitudinal histories of Leg 144 Guyots. *Proc. ODP B Sci. Res.* **144**, 915–933 (1995).
20. Johnson, C. C. *et al.* Middle Cretaceous reef collapse linked to ocean heat transport. *Geology* **24**, 376–380 (1996).
21. Skelton, P. W., Donovan, S. K. & Johnson, C. C. Middle Cretaceous reef collapse linked to ocean heat transport: Comment and Reply. *Geology* **25**, 477–479 (1997).
22. Lonsdale, P., Normark, W. R. & Neuman, W. A. Sedimentation and erosion on Horizon Guyot. *Bull. Geol. Soc. Am.* **83**, 289–315 (1972).
23. Saller, A. H. & Koepnick, R. B. Eocene to early Miocene growth of Enewetak Atoll: Insight from strontium-isotope data. *Bull. Geol. Soc. Am.* **102**, 381–390 (1990).
24. Land, L. A., Paull, C. K. & Hobson, B. Genesis of a submarine sinkhole without subaerial exposure: Straits of Florida. *Geology* **23**, 949–951 (1995).
25. Melim, L. A., Swart, P. K. & Maliva, R. G. Meteoric-like fabrics forming in marine waters; implications for the use of petrography to identify diagenetic environments. *Geology* **23**, 755–758 (1995).
26. Erba, E. *et al.* Synthesis of stratigraphies from shallow-water sequences at Sites 872 through 879 in the western Pacific Ocean (Leg 144). *Proc. ODP B Sci. Res.* **144**, 873–885 (1995).
27. Ludwig, K. R., Halley, R. B., Simmons, K. R. & Petermann, Z. E. Strontium-isotope stratigraphy of Enewetak Atoll. *Geology* **16**, 173–177 (1988).
28. Quinn, T. M., Lohmann, K. C. & Halliday, A. N. Sr isotopic variations in shallow water carbonate sequences: stratigraphic, chronostratigraphic, and eustatic implications of the record at Enewetak Atoll. *Paleoceanography* **6**, 371–385 (1991).
29. Aharon, P., Goldstein, S. L., Wheeler, C. W. & Jacobson, G. Sea-level events in the South Pacific linked with the Messinian salinity crisis. *Geology* **21**, 771–775 (1993).
30. Jenkyns, H. C., Paull, C. K., Cummins, D. & Fullagar, P. D. Strontium-isotope stratigraphy of Lower Cretaceous atoll carbonates in the Mid-Pacific Mountains. *Proc. ODP B Sci. Res.* **143**, 99–104 (1995).
31. Wilson, P. A., Opdyke, B. N. & Elderfield, H. E. Strontium isotope geochemistry of carbonates from Pacific guyots. *Proc. ODP B Sci. Res.* **144**, 447–457 (1995).
32. Wyatt, J. L., Quinn, T. M. & Davies, G. R. Petrology and geochemistry of limestones at Limalok and Wodejebato Guyots (Sites 871 and 874), Republic of the Marshall Islands. *Proc. ODP B Sci. Res.* **144**, 429–437 (1995).
33. McArthur, J. M., Thirlwall, M. F., Chen, M., Gale, A. S. & Kennedy, W. J. Strontium isotope stratigraphy in the Late Cretaceous: numerical calibration of the Sr isotope curve and intercontinental correlation for the Campanian. *Paleoceanography* **8**, 859–873 (1993).
34. Jones, C. E., Jenkyns, H. C., Coe, A. L. & Hesselbo, S. P. Strontium isotopic variations in Jurassic and Cretaceous seawater. *Geochim. Cosmochim. Acta* **58**, 3061–3074 (1994).
35. Bralower, T. J., Fullagar, P. D., Paull, C. K., Dwyer, G. S. & Leckie, R. M. Mid-Cretaceous strontium-isotope stratigraphy of deep-sea sections. *Bull. Geol. Soc. Am.* **109**, 1421–1442 (1997).
36. Wilson, P. A. & Opdyke, B. N. Equatorial sea-surface temperatures for the Maastrichtian revealed through remarkable preservation of metastable carbonate. *Geology* **24**, 555–558 (1996).
37. Wilson, P. A. & Dickson, J. A. D. Radial calcite: Alteration product of and petrographic proxy for magnesium calcite marine cement. *Geology* **24**, 945–948 (1996).
38. Gross, M. G. & Tracey, J. I. Oxygen and carbon isotopic composition of limestones and dolomites, Bikini and Eniwetok Atolls. *Science* **151**, 1082–1084 (1966).
39. Aharon, P. Analysis of the anomalous $^{18}\text{O}/^{16}\text{O}$ and D/H isotope ratios in tropical rainfall over the western Pacific Ocean. *Eos* **64**, 106 (1983).
40. Saller, A. H. & Moore, C. H. Geochemistry of meteoric calcite cements in some Pleistocene limestones. *Sedimentology* **38**, 601–621 (1991).
41. Douglas, R. G. & Savin, S. M. Oxygen and carbon isotope analyses of Cretaceous and Tertiary foraminifera from the central North Pacific. *Init. Rep. DSDP* **17**, 591–605 (1973).
42. Gradstein, F. M. *et al.* A Mesozoic time scale. *J. Geophys. Res.* **99**, 24051–24074 (1994).
43. Arnaud Vanneau, A. & Premoli Silva, I. Biostratigraphy and systematic description of benthic foraminifera from mid-Cretaceous shallow-water carbonate platforms at Sites 878 and 879 (MIT and Takuyo-Daisan Guyots). *Proc. ODP B Sci. Res.* **144**, 199–214 (1995).
44. Tarduno, J. A. & Gee, J. Large-scale motion between Pacific and Atlantic hotspots. *Nature* **378**, 477–480 (1995).
45. Sager, W. W. & Pringle, M. S. Mid-Cretaceous to early Tertiary apparent polar wander path of the Pacific Plate. *J. Geophys. Res.* **93**, 11753–11771 (1988).
46. Larson, R. L. & Sager, W. W. Skewness of magnetic anomalies M0 to M29 in the northwestern Pacific. *Proc. ODP B Sci. Res.* **129**, 471–481 (1992).
47. Lincoln, J. M., Pringle, M. S. & Premoli Silva, I. Early and Late Cretaceous volcanism and reef-building in the Marshall Islands. *Geophys. Monogr.* **77**, 279–305 (1993).
48. Spencer, T. Rates of karst processes on raised reef limestones and their implications for coral reef histories. *Coral Reef Cong. Proc.* **5**, 629–634 (1985).
49. Murray, J. W., Barber, R. T., Roman, M. R., Bacon, M. P. & Feely, R. A. Physical and biological controls on carbon cycling in the Equatorial Pacific. *Science* **266**, 58–65 (1994).
50. Jenkyns, H. C. Carbon-isotope stratigraphy and paleoceanographic significance of the Lower Cretaceous shallow-water carbonates of Resolution Guyot, Mid-Pacific Mountains. *Proc. ODP B Sci. Res.* **143**, 99–108 (1995).
51. Brown, B. E. Coral bleaching: Causes and consequences. *Coral Reefs* **16**, 129–138 (1997).
52. Tarduno, J. A. & Sager, W. W. Polar standstill of the Mid-Cretaceous Pacific Plate and its geodynamic implications. *Science* **269**, 956–959 (1995).

Acknowledgements. We thank our colleagues from ODP Legs 143 and 144. In particular, we acknowledge the critical collaboration of E. Erba in the original development (ref. 19) of the 'Death in the Tropics' hypothesis and B. N. Opdyke for support and encouragement of a pilot project. We thank M. Greaves and E. Hawkins (Cambridge), J. Cartledge (Oxford) and Lora Wingate (Michigan) for laboratory assistance. This work was supported by UK Ocean Drilling Program with continuing support from a UK National Environmental Research Council Fellowship.

Correspondence and requests for materials should be addressed to P.A.W. (e-mail: paw1000@cam.ac.uk).

Vigorous star formation hidden by dust in a galaxy at a redshift of 1.4

Andrea Cimatti*, Paola Andreani†, Huub Röttgering‡ & Remo Tilanus§||

* Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125, Firenze, Italy

† Dipartimento di Astronomia, vicolo dell'Osservatorio 5, I-35122 Padova, Italy

‡ Sterrewacht, Postbus 9513, Leiden, 2300 RA, The Netherlands

§ Joint Astronomy Centre, 660 N. A'ohoku Place, Hilo, Hawaii 96720, USA

|| Netherlands Foundation for Research in Astronomy, Dwingeloo, The Netherlands

Near-infrared surveys have revealed a substantial population of faint galaxies that are extremely red in their optical to near-infrared colours¹. These enigmatic galaxies are about as numerous as faint quasars¹. The extremely red colours could arise either because of an absence of recently formed stars, implying that the galaxies formed almost all of their stars at very high redshifts², or because the galaxies contain large amounts of dust, which reddens the light by preferentially absorbing the shorter-wavelength radiation. Determining why these objects are red will help us to understand the early evolution of galaxies. The object HR10, at a redshift of $z = 1.44$ (refs 1, 3), is considered the archetype of the extremely red galaxies. Here we report the detection of continuum emission from warm dust in HR10, demonstrating that it is a dusty galaxy undergoing a burst of massive-star formation. Our result provides a clear example of a high-redshift galaxy where the star-formation rate inferred from the ultraviolet luminosity (the usual method at these redshifts) would be underestimated by a factor of more than 500, showing that great caution must be used to infer the global star-formation history of the Universe from optical and ultraviolet observations⁴.

The current knowledge of the global star-formation activity in the Universe is mostly based on the ultraviolet (UV) continuum luminosity of high-redshift galaxies⁵. However, a crucial phase in the evolution of a galaxy might be a vigorous starburst producing large numbers of luminous stars. In this stage, most of the UV stellar radiation would be absorbed by the dust grains and re-emitted in the far-infrared (FIR), making the galaxy dark at UV and optical wavelengths. A cosmic background has been first detected at $\lambda_{\text{obs}} \approx 200 \mu\text{m}$ – 2 mm (ref. 6) and then interpreted⁴. In addition, a cosmic background at $140 \mu\text{m}$ and $240 \mu\text{m}$ has been recently found⁷. The presence of such a background suggests the existence of a population of distant dusty galaxies. Several models predict the number of dusty galaxies at high redshift (refs 4, 8–10). However, dusty galaxies at $z > 1$ that are known at present are limited to a handful of extremely rare objects such as radio galaxies, quasars and IRAS-selected luminous galaxies (see refs 11, 12 and references therein), and the existence of a general field population of dusty galaxies has not been established yet. Recent surveys for field galaxies found faint (R -band magnitude $R > 24$) and very red objects with colours typically in the range^{1,13} of $7 < (R - K) < 8$. Their sky surface density ~ 0.01 – 0.02 arcmin^{-2} at $K \leq 20$ (refs 1, 13), is comparable to that of faint quasars¹, but about one order of magnitude less than that of field galaxies with $z > 3$ (ref. 14).

In order to test if a fraction of these galaxies have these extremely red colours because of strong dust reddening, we started a programme aimed at detecting their dust thermal emission at submillimetre and millimetre wavelengths. Our first target was HR10, one of the reddest galaxies known to date, with $I = 24.9$, $R - K \approx 8$, $I - K = 6.5$ (ref. 1) and the only one with a measured³ redshift ($z = 1.44$). Its UV–optical spectral energy distribution suggested

the presence of dust reddening^{3,15}. HR10 was observed and detected with the Institut de Radioastronomie Millimétrique (IRAM) 30-m antenna at $1,250 \mu\text{m}$ as well as with the James Clerk Maxwell Telescope (JCMT) 15-m telescope at $850 \mu\text{m}$ in March and September 1997, respectively. The details of the observations and data reduction can be found in Fig. 1 legend. Here we assume $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $q_0 = 0.5$, and we define $h_{50} = H_0/50$.

Figure 1 shows the spectral energy distribution of HR10. Synchrotron radiation emission as the source of the submillimetre fluxes can be ruled out because the radio fluxes^{3,16} constrain the extrapolated synchrotron flux at $1,250 \mu\text{m}$ to be $< 0.34 \text{ mJy}$, more than an order of magnitude less than that observed. We therefore interpret the continuum emission observed at 850 and $1,250 \mu\text{m}$ as due to thermal dust emission. Using the ISO upper limit at $175 \mu\text{m}$ (ref. 17), and for dust emissivity indices¹⁸ $\beta = 1$ – 2 , the dust temperature T_d is found to be in a range $18 < T_d < 45 \text{ K}$. These temperatures fall into the range of those of active and star-forming galaxies (see refs 11, 12, 19 and references therein). The total dust mass can be estimated as²⁰ $M_d = S_{\nu_{\text{obs}}} D_L^2 / [(1 + z) \kappa_d(\nu_{\text{rest}}) B(\nu_{\text{rest}}, T_d)]$, where ν_{obs} and ν_{rest} are, respectively, the observed and rest-frame frequencies, $S_{\nu_{\text{obs}}}$ is the observed flux density at ν_{obs} , D_L is the luminosity distance, B is the black-body Planck function, T_d is the dust temperature and $\kappa_d = 0.67(\nu_{\text{rest}}/250 \text{ GHz})^\beta \text{ cm}^2 \text{ g}^{-1}$ is the adopted mass absorption coefficient¹¹. In the two extreme cases considered in Fig. 1, the total dust mass is $M_d = (7.3 \times 10^8) h_{50}^{-2} M_\odot$ and $M_d = (3.3 \times 10^9) h_{50}^{-2} M_\odot$ for $T_d = 45 \text{ K}$ and $T_d = 18 \text{ K}$, respectively. ($M_\odot$ is the solar mass.) The dust mass increases or decreases by a factor of ~ 2 if $\beta = 1$ or $\beta = 2$, respectively. If we assume²¹ $M_{\text{H}_2}/M_d = 100$ the mass of molecular hydrogen is of the order of $M_{\text{H}_2} \approx 10^{11} h_{50}^{-2} M_\odot$. The total FIR luminosity is estimated by integrating the grey-body curves in the range $\lambda_{\text{rest}} = 10$ – $2,000 \mu\text{m}$, $L_{\text{FIR,rest}} = 4\pi D_L^2 (1 + z)^2 \int S_{\nu_{\text{obs}}} d\nu$, $S_{\nu_{\text{obs}}}$ is the flux density. We find that $L_{\text{FIR,rest}} = (3.8 \times 10^{12}) h_{50}^{-2} L_\odot$ and $L_{\text{FIR,rest}} = (1.5 \times 10^{13}) h_{50}^{-2} L_\odot$ for $T_d = 18 \text{ K}$ and $T_d = 45 \text{ K}$, respectively ($L_\odot$ is the solar luminosity).

At low redshifts, HR10 is comparable to the ultra-luminous infrared galaxies which are dusty systems with $L_{\text{FIR}} > 10^{12} L_\odot$ and typically $T_d \approx 30$ – 60 K (ref. 12), but with dust masses $M_d \approx (10^7$ – $10^8) M_\odot$ lower than those inferred for HR10 (see ref. 12 and references therein). At high redshifts, HR10 can be compared to IRAS10214+4724, an ultra-luminous infrared galaxy at $z = 2.286$ (ref. 22) whose properties are severely affected by gravitational lensing. If the correction for lensing magnification is taken into account ($\sim 30\times$ in the infrared²³), IRAS10214+4724 has $L_{\text{FIR}} = (4.4 \times 10^{12}) h_{50}^{-2} L_\odot$ and $M_d = (1.1 \times 10^8) h_{50}^{-2} M_\odot$, comparable to those of HR10, but it has a warmer dust temperature ($40 < T_d < 80 \text{ K}$; see refs 11, 12 and references therein). The warmer dust could be due to the additional source of UV radiation provided by the hidden quasar present in the nucleus of IRAS10214+4724 (ref. 24).

If we apply the relation between the star formation rate and the FIR luminosity: $\text{SFR} = (\Psi 10^{-10} L_{\text{FIR}}/L_\odot) M_\odot \text{ yr}^{-1}$, where $\Psi = 0.8$ – 2.1 (see ref. 25 and references therein), and we adopt $\Psi = 1.5$, we derive $\text{SFR} \approx (570$ – $2,250) h_{50}^{-2} M_\odot \text{ yr}^{-1}$, which is larger than the typical SFRs of low-redshift ultra-luminous infrared galaxies ($\sim (10$ – $100) M_\odot \text{ yr}^{-1}$ (ref. 12)), but comparable to that of IRAS10214+4724 ($\sim 660 M_\odot \text{ yr}^{-1}$ (ref. 11)). If HR10 contains an obscured active galactic nucleus heating the dust, then only a fraction of L_{FIR} can be ascribed to star formation. The H α line detected in a very low-resolution spectrum of HR10 (ref. 3) has a signal-to-noise ratio insufficient to establish the presence of a broad component of this line, which would imply that the nucleus of the quasar is not severely obscured. However, the K-band morphology of HR10 is spiral-like and not strongly nucleated, and it does not support the presence of a directly visible quasar nucleus.

Owing to the steep rise of the grey-body dust spectra towards $\lambda_{\text{rest}} \approx 100$ – $200 \mu\text{m}$, the observations at longer wavelengths make

the submillimetre and millimetre flux (at a fixed observed λ) of a dusty galaxy roughly constant for approximately $1 < z < 10$ (refs 8, 9, 11). For instance, we assumed that HR10 had three different redshifts ($z = 0.7, 2.2, 3.3$) and we computed how its properties would change accordingly, adopting $\beta = 1.5$ and an optically thin grey-body: $S_\nu \propto \nu^\beta B_\nu(T)$. We find that the inferred dust temperature would increase from ~ 15 – 20 K at $z = 0.7$ to ~ 45 – 50 K at $z = 3.3$. L_{FIR} and M_d would be $\sim 0.1, 7, 25 \times 10^{13} h_{50}^{-2} L_\odot$ and $\sim 30, 10, 5.6 \times 10^8 h_{50}^{-2} M_\odot$ for $z = 0.7$ ($T_d = 18$ K), $z = 2.2$ ($T_d = 35$ K), and $z = 3.3$ ($T_d = 45$ K), respectively. The main implication is that a galaxy with observed submillimetre and millimetre properties as those of HR10 would be classified a dusty ultra-luminous infrared galaxy in a wide range of redshifts.

The discovery of dust associated with HR10 has several implications. First, it sheds light on the nature of this galaxy, showing that HR10 is a very dusty star-forming galaxy where, similarly to low-

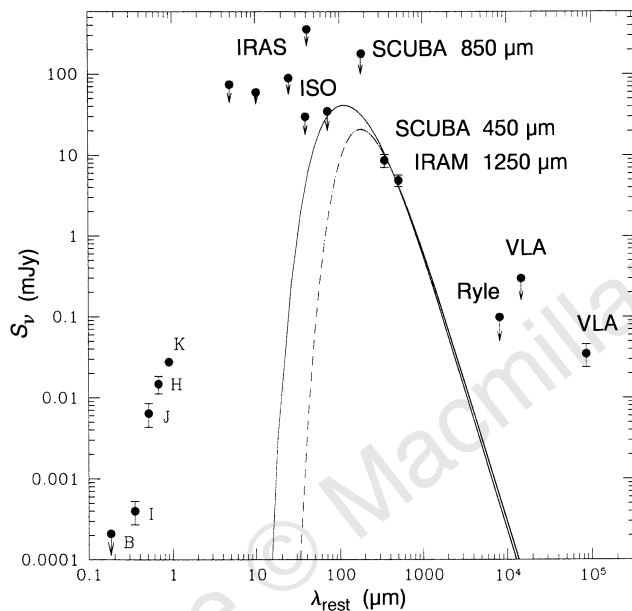


Figure 1 Spectral energy distribution of HR10. Filled circles, fluxes of HR10 taken from refs 1, 3, 16, 17 and this work. The Infrared Astronomical Satellite (IRAS) 3 σ upper limits at $\lambda_{\text{obs}} = 12, 25, 60, 100 \mu\text{m}$ were derived from co-added survey data with the SCANPI procedure available at IPAC (Infrared Processing and Analysis Center in Pasadena, California, USA). The curves show two extreme cases of grey-bodies consistent with the data. Dashed curve, optically thin grey-body, $S_\nu \propto \nu^\beta B_\nu(T)$, with $\beta = 1.5$ and $T_d = 18$ K. Continuous curve, grey-body optically thin at $\nu < \nu_0 = 750$ GHz, $S_\nu \propto B_\nu(T) \{1 - \exp[-(\nu/\nu_0)^\beta]\}$ (ref. 11), with $\beta = 1.5$ and $T_d = 45$ K. HR10 was observed for 64 min (on source) with the Max-Planck-Institut für Radioastronomie 19-channel bolometer²⁹ (1,250 μm , 11" (full-width at half-maximum) beam, 32" chop throw, zenith opacity 0.09–0.3). Flux calibration was achieved using Uranus as primary calibrator and Mars and pointing quasars as secondary ones. The mean atmospheric level was found by averaging the measurements of 16 channels and subtracted from the central pixel, which targeted HR10. The final flux at 1,250 μm is 4.9 ± 0.7 mJy. The simultaneous observations at 450 and 850 μm were made with the Submillimetre Common User Bolometer Array (SCUBA) (E. I. Robson *et al.*, manuscript in preparation) at the JCMT in photometry mode (108 min on source, zenith opacity 0.3 at 850 μm). The sky background was removed from the flat-fielded and despiked data by subtracting the average signal from five-neighbour bolometers (containing no source signal). The data were calibrated using photometry and beam maps of Mars and Uranus. The flux at 850 and 450 μm was found to be 8.7 ± 1.6 mJy and < 180 mJy (3 σ) respectively. The 1 σ error bars of the 850 and 1,250 μm flux densities indicate the statistical uncertainties, whereas the uncertainty on the absolute flux calibration is estimated to be ~ 15 – 20% both in the IRAM and SCUBA data. ISO, Infrared Space Observatory; VLA, Very Large Array; Ryle, Ryle Telescope.

redshift ultra-luminous infrared galaxies, most of the energy is emitted in the far-infrared. The second implication is generally related to the problem of estimating the star-formation rate in distant galaxies. The global history of star formation in the Universe is at present inferred mostly by the continuum UV luminosity of optically selected high-redshift galaxies⁵, from which the derived SFRs are in the range of $(4\text{--}25)h_{50}^{-2} M_\odot \text{yr}^{-1}$ (ref. 14). However, such estimates can be severely hampered by the presence of dust extinction which modifies the shape and reduces the flux of the UV spectra. Observations of low-redshift galaxies have also pointed out the limits of the UV-blue light as a SFR estimator (see, for instance, ref. 26 and references therein). HR10 allows us to see an extreme example of this problem applied to a high-redshift system. In fact, the H α luminosity of HR10 would imply $\text{SFR} \approx 80h_{50}^{-2} M_\odot \text{yr}^{-1}$ (ref. 3), whereas the luminosity of the UV continuum at 2,800 Å would suggest $\text{SFR} \approx 1h_{50}^{-2} M_\odot \text{yr}^{-1}$ (ref. 5), both at least one order of magnitude less than the SFR suggested by the FIR luminosity. In this regard, our results show that the global star formation history of the Universe can be fully traced only if the effects of dust are taken into account. Objects like HR10 would not be found by optical imaging surveys based either on the Lyman-continuum break¹⁴ or on strong emission lines; neither would they have been found by IRAS surveys or by traditional quasar surveys. Instead, our results show that the combination of optical and near-infrared deep imaging, coupled with (sub)millimetre observations, is an efficient method to find dusty galaxies at high-redshift. Indeed, recent deep imaging using the SCUBA bolometer array (see Fig. 1 legend) suggests the presence of a population of faint sub-millimetre sources²⁷.

Our observations of HR10 suggest that star formation in distant objects occurs in different modes, and that the most vigorous episodes of star formation probably arise in dusty environments as predicted by several models^{9,10,28}. However, our data cannot tell us if HR10 is forming its first generation of stars, or if the starburst is occurring in an already formed system. Nevertheless, in both cases we are witnessing a significant episode of star formation: if the burst lasts $\sim 10^7$ yr (a typical timescale for a starburst¹²), the total mass of gas converted into stars is of the order of $10^{10} M_\odot$, which is $\sim 10\%$ of the total mass of a present-day massive galaxy.

We note that the observed properties of HR10 fit predictions⁴ that the sources at $z \approx 0.5$ – 2.5 which contribute to the cosmic FIR–millimetre background should have fluxes of around 10–100 mJy at $\lambda_{\text{obs}} = 200 \mu\text{m}$. For $18 \leq T_d \leq 45$ K, the expected flux of HR10 at $\lambda_{\text{obs}} = 200 \mu\text{m}$ is in the range 10–40 mJy. Moreover, the predicted sky surface density of dusty galaxies at $\lambda_{\text{obs}} = 175 \mu\text{m}$ ($\sim 0.05 \text{ arcmin}^{-2}$) is (within a factor of five) similar to that of the extremely red galaxies^{1,13}. Further observations would provide clues on what fraction of the extremely red galaxies contributes to the FIR–millimetre background. □

Received 27 October 1997; accepted 4 March 1998.

- Hu, E. M. & Ridgway, S. E. Two extremely red galaxies. *Astron. J.* **107**, 1303–1306 (1994).
- Dunlop, J. *et al.* A 3.5-Gyr-old galaxy at redshift 1.55. *Nature* **381**, 581–584 (1996).
- Graham, J. R. & Dey, A. The redshift of an extremely red object and the nature of the very red galaxy population. *Astrophys. J.* **471**, 720–725 (1996).
- Guideroni, B. *et al.* The optically dark side of galaxy formation. *Nature* **390**, 257–259 (1997).
- Madau, P., Pozzetti, L. & Dickinson, M. The star formation history of field galaxies. *Astrophys. J.* (in the press).
- Puget, J. L. *et al.* Tentative detection of a cosmic far-infrared background with COBE. *Astron. Astrophys.* **308**, L5–L8 (1996).
- Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust IR emission for use in estimation of reddening and CMBR foregrounds. *Astrophys. J.* (in the press).
- Blain, A. W. & Longair, M. S. Millimetre background radiation and galaxy formation. *Mon. Not. R. Astron. Soc.* **265**, L21–L24 (1993).
- Franceschini, A., Mazzei, P., Danese, L. & De Zotti, G. Luminosity evolution and dust effects in distant galaxies: implications for the observability of the early evolutionary phases. *Astrophys. J.* **427**, 140–154 (1994).
- Zepf, S. E. & Silk, J. On the effects of bursts of massive star formation during the evolution of elliptical galaxies. *Astrophys. J.* **466**, 114–121 (1996).
- Hughes, D. H., Dunlop, J. S. & Rawlings, S. High-redshift radio galaxies and quasars at submillimetre wavelengths: assessing their evolutionary status. *Mon. Not. R. Astron. Soc.* **289**, 766–782 (1997).
- Sanders, D. B. & Mirabel, I. F. Luminous infrared galaxies. *Annu. Rev. Astron. Astrophys.* **34**, 749–782 (1996).
- Thompson, D. & Beckwith, S. V. W. In *The Young Universe: Galaxy Formation and Evolution at*

- Intermediate and High Redshift* (eds. D'Odorico, S., Fontana, A. & Giallongo, E.) (ASP, Conf. Ser., Astron. Soc. Pacific, in the press).
14. Steidel, C. C. *et al.* Spectroscopic confirmation of a population of normal star-forming galaxies at redshifts $z > 3$. *Astrophys. J.* **462**, L17–L21 (1996).
 15. Cimatti, A., Bianchi, S., Ferrara, A. & Giovanardi, C. On the dust extinction in high- z galaxies and the case of extremely red objects. *Mon. Not. R. Astron. Soc.* **290**, L43–L49 (1997).
 16. Jones, M. E. *et al.* Detection of a cosmic microwave background decrement toward the $z = 3.8$ quasar pair PC 16434631A,B. *Astrophys. J.* **479**, L1–L3 (1997).
 17. Ivison, R. J., Archibald, E. N., Dey, A. & Graham, J. R. in *The Far-Infrared and Submillimetre Universe* (ed. Wilson, A.) 281–284 (SP-401, ESA Publ. Div., ESTEC, Noordwijk, 1998).
 18. Whittet, D. C. B. *Dust in the Galactic Environment* (Inst. Physics, Bristol, 1992).
 19. Andreani, P. & Franceschini, A. 1.25-mm observations of a complete sample of IRAS galaxies. II. Dust properties. *Mon. Not. R. Astron. Soc.* **283**, 85–100 (1997).
 20. Hildebrand, R. H. The determination of cloud masses and dust characteristics from submillimetre thermal emission. *Q.J.R. Astron. Soc.* **24**, 267–282 (1983).
 21. Solomon, P. M., Downes, D., Radford, S. J. E. & Barrett, J. W. The molecular interstellar medium in ultraluminous infrared galaxies. *Astrophys. J.* **478**, 144–161 (1997).
 22. Rowan-Robinson, M. *et al.* The ultraviolet-to-radio continuum of the ultraluminous galaxy IRAS F10214+4724. *Mon. Not. R. Astron. Soc.* **261**, 513–521 (1993).
 23. Eisenhardt, P. R. *et al.* Hubble space telescope observations of the luminous IRAS source FSC 10214+4724: A gravitationally lensed infrared quasar. *Astrophys. J.* **461**, 72–83 (1996).
 24. Goodrich, R. W. *et al.* FSC 10214+4724: a gravitationally lensed, hidden QSO. *Astrophys. J.* **456**, L19–L22 (1996).
 25. Rowan-Robinson, M. *et al.* Observations of the Hubble Deep Field with the Infrared Space Observatory—V. Spectral energy distributions, starburst models and star formation history. *Mon. Not. R. Astron. Soc.* **289**, 490–496 (1997).
 26. Gallagher, J. S. III, Hunter, D. A. *Star Formation in Galaxies* (ed. Lonsdale-Persson, C. J.) 167–177 (CP-2466, NASA, Pasadena, 1987).
 27. Smail, I., Ivison, R. J. & Blain, A. W. A deep sub-millimetre survey of lensing clusters: a new window on galaxy formation and evolution. *Astrophys. J.* (in the press).
 28. Zepf, S. E. Formation of elliptical galaxies at moderate redshifts. *Nature* **390**, 377–379 (1997).
 29. Kreysa, E. in *Proc. Int. Symp. on Photon Detectors for Space Instrumentation* (ed. Guyenne, T. D.) 45–48 (ESA/ESTEC, Noordwijk, 1993).

Acknowledgements. We thank S. Randich, D. Hughes, R. Ivison, M. Lehnert and J. Baker for discussions, and E. Hu for checking the astrometry of HR10 and for suggestions. The JCMT is operated by the Joint Astronomy Centre on behalf of the UK PPARC, the Netherlands Organisation for Scientific Research, and the National Research Council of Canada.

Correspondence and requests for materials should be addressed to A.C. (e-mail: cimatti@arcetri.astro.it).

Discovery of two distant irregular moons of Uranus

Brett J. Gladman*, Philip D. Nicholson†, Joseph A. Burns†, J J Kavelaars‡, Brian G. Marsden§, Gareth V. Williams§ & Warren B. Offutt||

* Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, Ontario M5S 3H8, Canada

† Department of Astronomy, Cornell University, Ithaca, New York 14853-6801, USA

‡ Department of Physics and Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada

§ Smithsonian Astrophysical Observatory, Cambridge, Massachusetts 02138, USA

|| W & B Observatory, PO Drawer 1130, Cloudcroft, New Mexico 88317, USA

The systems of satellites and rings surrounding the giant planets in the Solar System have remarkably similar architectures¹. Closest to each planet are rings with associated moonlets, then larger 'regular' satellites on nearly circular orbits close to the planet's equatorial plane, and finally one or more distant, small 'irregular' satellites on highly elliptical or inclined orbits. Hitherto, the only departure from this broad classification scheme was the satellite system around Uranus, in which no irregular satellites had been found². Here we report the discovery of two satellites orbiting Uranus at distances of several hundred planetary radii. These satellites have inclined, retrograde orbits of moderate eccentricity that clearly identify them as irregular. The satellites are extremely faint (apparent red magnitudes $m_R = 20.4$ and 21.9), with estimated radii of only 60 and 30 km. Both moons are unusually red in colour, suggesting a link between these objects—which were presumably captured by Uranus early in the Solar System's history—and other recently discovered bodies³ orbiting in the outer Solar System.

Jupiter has eight known irregular satellites, of which the last, Leda, was discovered by Kowal and co-workers in 1974⁴. Saturn and

Neptune have one apiece, Phoebe being spotted by Pickering in 1898⁵ while Nereid was found by Kuiper in 1949⁶. A summary of data on the irregular satellites and a discussion of their likely origins as captured planetesimals are given in ref. 7.

All 15 previously known uranian satellites lie on fairly evenly spaced, nearly circular orbits of low inclination with respect to the planet's equator, which is tilted at 98° relative to the planet's orbital plane. The largest and outermost regular satellites, Titania and Oberon, were discovered by Herschel in 1787, 6 years after he found Uranus, while Lassell discovered Ariel and Umbriel in 1851. Miranda was first seen by Kuiper⁸ in 1948. The orbital radii of these satellites range from 5.07 to $22.77 R_U$, where Uranus's standard equatorial radius R_U is 25,559 km. In 1985–86 Voyager 2 discovered 10 small, dark, inner moons, clustered between 1.94 and $3.36 R_U$ (ref. 9).

Several previous searches have been made for faint uranian satellites. From photographic plates taken on the Mount Wilson 60-inch reflector, Christie¹⁰ concluded that no new object brighter than 19th magnitude lay within $10'$ of Uranus. (Note that $1'$ corresponds to $\sim 790,000$ km or $31 R_U$ at Uranus's mean opposition distance). Using plates from the McDonald 82-inch reflector,

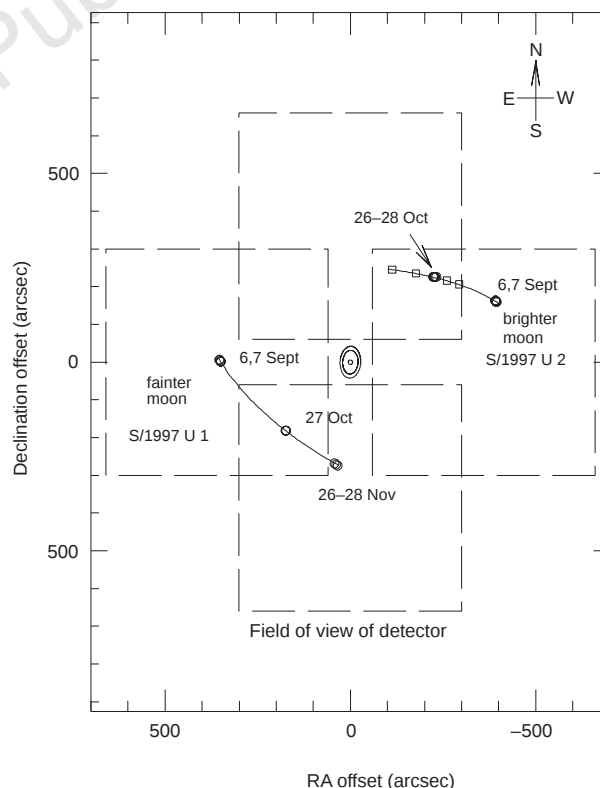


Figure 1 Observed positions of the newly discovered moons relative to Uranus. At the centre is the planet (not to scale), surrounded by the circular orbits of Titania and Oberon, inclined by an angle of 41° to the line of sight (RA, right ascension). Arcs followed by the satellites from their discovery until 28 November 1997 are shown, corresponding to preliminary orbit fits (Table 1). Palomar observations in September and October, as well as the CFHT observations in November, are shown as open circles, while squares indicate selected observations from Cloudcroft. Dashed boxes show the four CCD fields used for the searches of 6 and 7 September 1997 (UT) at Palomar, in which the moons were later discovered. The 300-s exposures, taken through a Gunn r filter (610–700 nm), have limiting magnitudes (5σ) of $m_R = 23.5$; seeing was $\sim 1.3''$ full-width at half-maximum. The detector was a $2,048 \times 2,048$ thinned Tektronix CCD chip with a quantum efficiency of 85–90% between 550 and 750 nm and a fast readout (40 s with 2×2 binned $0.56''$ pixels).

Kuiper⁸ set a limiting photographic magnitude of 20.5 on new moons within $\sim 1^\circ$ of the planet. Attempting to identify faint inner satellites before the Voyager encounter, Smith¹¹, observing with R. Terrile, masked the planet with a coronagraph and searched in the deep methane absorption band at 890 nm to a limiting magnitude of 23, but only out to $\sim 2'$ from the planet. In 1984 Cruikshank (ref. 12 and personal communication) obtained several deep photographic exposures with the 3.6-m Canada–France–Hawaii Telescope's (CFHT) prime-focus camera, using an orange filter transmitting wavelengths greater than 570 nm, and surveyed 90% of the volume out to $900 R_U$ down to $m_V \approx 22.5$ without success. Most recently, Żytkow *et al.*¹³ unsuccessfully searched a $3^\circ \times 3^\circ$ region centred on the planet by the automated scanning of blue-sensitive plates taken with the UK Southern Schmidt Telescope in 1990, estimating 95% completeness for satellites brighter than $B_J = 20.6$. Although cameras aboard the Voyager 2 spacecraft scanned the plane's immediate vicinity for faint inner moons, the restricted field of view of the narrow-angle camera (7×10^{-3} rad), combined with its limited sensitivity, rendered impractical a survey of the regions exterior to Oberon (C. Porco, S. Synnott and R. Terrile, personal communications).

The size of stable prograde satellite orbits is limited by the radius of the Hill sphere¹⁴, the region wherein the planet's gravity exceeds the sun's tidal force. For Uranus this implies a maximum projected radius of $2,800 R_U$, or $\sim 1.5^\circ$ on the sky. Retrograde orbits are stable to somewhat greater distances¹⁵. Although the area covered by our CCD field ($9.7' \times 9.7'$) was small compared to the Hill sphere, all known irregular satellites orbit within 350 planetary radii of their primaries, or $11'$ at Uranus. This fact, and the knowledge that we could reach ~ 2 mag fainter than previous upper limits, encouraged us to conduct a brief search for satellites over several hours available during another observing programme in September 1997.

In contrast to regular satellites, which lie close to their planet's equatorial plane, distant moons are dominated by solar perturbations and maintain roughly constant inclinations with respect to the planet's orbital plane¹⁶. For Uranus, solar perturbations become dominant outside $108 R_U$, accounting for secular perturbations from the inner satellites. However, many irregular satellites have large inclinations relative to their planet's orbital plane¹⁷, so that there is no preferred plane in which to search. We thus chose four partially overlapping fields surrounding Uranus covering distances from $\sim 1'$ to $11'$ from the planet (see Fig. 1). The inner limit was set to avoid excess scattered light from Uranus itself.

We used the Cosmic CCD camera mounted at the prime focus of the 5-m Hale Telescope at Palomar Observatory, with a red filter to suppress night sky background and exploit the peak quantum efficiency of the detector. Three exposures of each search field, spaced ~ 1 h apart, were taken on each of two successive nights. Uranus's westward motion with respect to the background stars on these nights was $4.2'' \text{ h}^{-1}$, providing an easily detectable displacement between frames. Because the circumplanetary motion for a body orbiting Uranus at a mean radius of $6'$ is at most $0.28'' \text{ h}^{-1}$ (scaling as the inverse square root of the separation), the orbital motions of any satellites within our search area should be negligible compared to Uranus's rate. Such moons should thus 'follow the planet' across the sky over a single night's observations. Main-belt asteroids near the stationary points of their retrograde loops have

slow right ascension motions that, at Uranus's 40° elongation from opposition in September, can nearly match the planet's. Fortunately the declination motions of such asteroids are generally much greater than that of Uranus, which allowed us to reject several such asteroids in our search.

The images were examined in early October to identify slowly moving objects, using techniques employed to search for Kuiper-belt objects¹⁷. An object, with R-band magnitude ~ 21.9 and designated S/1997 U 1 (referred to here as U 1), was detected $6'$ almost due east of Uranus, moving slowly southwest relative to the planet. A much brighter object (S/1997 U 2, referred to here as U 2, with $m_R \approx 20.4$) was found $6.5'$ west and $2.5'$ north of Uranus, and moving northeast relative to the planet. The measured relative motions of $0.22 \pm 0.03'' \text{ h}^{-1}$ and $0.13 \pm 0.03'' \text{ h}^{-1}$ for U 1 and U 2, respectively, were roughly consistent with circular planetocentric orbital motion at their projected distances.

U 1 was measured in five frames over the two nights, and U 2 in all six available frames. The U 2 discovery images are shown in ref. 18. Astrometric positions for both objects were calculated by computing plate solutions using reference-star coordinates obtained from the PMM USNO-A1.0 catalogue. Initial heliocentric orbit fits indicated that both objects were almost certainly at distances comparable to that of Uranus. Although the possibility existed that these were objects in heliocentric orbits fortuitously passing near Uranus, the low sky density of Centaurs (~ 0.01 per square degree brighter than $m_R = 23.0$; ref. 3) made this extremely unlikely; the probability that two such objects would be detected within our search area of ~ 0.1 square degrees is $\sim 10^{-6}$.

The brighter object was observed on eight nights in October using W.O.'s CCD-equipped 0.6-m telescope at Cloudcroft, New Mexico. Detections were rendered difficult by a full Moon on 16 October and the proximity of Uranus to its stationary point on 15 October, but the measured positions strengthened the case for circumplanetary versus heliocentric motion. On 27 October, both objects were re-observed at Palomar, with several images being obtained through B and I filters to provide colour information. Poor seeing on 26 October and thin clouds on 28 October prevented detection of U 1 on these nights, but additional images of U 2 were secured. D. J. Tholen (assisted by C. Herrick and R. J. Whitely) detected both objects on 29 October with the University of Hawaii's 2.2-m reflector, as did A. Fitzsimmons and M. E. Fletcher at the 2.5-m Isaac Newton Telescope on La Palma on 30 October. S. Lilly and co-workers observed U 1 using the CFHT on 26–28 November, greatly improving its orbit determination. Selected astrometric positions, including some of U 2 obtained in early November by C. W. Hergenrother with the Smithsonian Astrophysical Observatory's 1.2-m reflector in Arizona as well as additional observations from Cloudcroft in late November (Fig. 1), are given in refs 19–21. Most recently we have identified two objects on the CFHT plates taken by D. P. Cruikshank in June 1984 that may be U 1 and U 2; further analysis is continuing.

By late October, orbital computations suggested that the observations of U 2 were incompatible with a heliocentric orbit, although such a statement could not yet be made for U 1 because its observations consisted of essentially only two points. However, a circumplanetary orbit was still favoured for the statistical reason mentioned above. Although ambiguity exists in the orientation of

Table 1 Data for the newly discovered uranian satellites

Object	Magnitudes			<i>r</i> (km)	<i>P</i> (days)	<i>a</i> (R_U)	<i>e</i>	<i>i</i> (deg.)	Ω (deg.)	ω (deg.)
	m_R	<i>B</i> – <i>R</i>	<i>R</i> – <i>I</i>							
S/1997 U 1	21.9 ± 0.1	1.5 ± 0.2	0.5 ± 0.2	30	654	305	0.2	146	186	142
S/1997 U 2	20.4 ± 0.1	1.6 ± 0.2	0.6 ± 0.2	60	495	253	0.4	153	221	256

The apparent magnitudes are in the Kron–Cousins system. Radii *r* are for assumed geometric albedos of 0.07. A full set of planetocentric orbital elements is given in ref. 21; inclinations *i* are with respect to the ecliptic of J2000. For U 2, uncertainties are ~ 0.2 in the eccentricity *e*, $\sim 5^\circ$ in inclination *i* and $\sim 45^\circ$ in the node Ω . The argument of pericentre ω , semimajor axis *a* and orbital period *P* are much less certain, as are all the elements for U 1.

any orbit observed over a very limited range of geometries, it appeared likely that the orbit of U 2 was retrograde with respect to Uranus's orbital motion, inclined by $\sim 30^\circ$ relative to the ecliptic, and quite eccentric. The orbital period was ~ 415 d with semimajor axis $219 R_U$. The near-circular orbit computed²² for U 1 at this time was arbitrarily assumed to be retrograde and with the same semimajor axis as U 2. The late-November observations showed U 1 to be less than $15''$ from this arbitrary prediction, a discordance easily removed by a minor adjustment. Unpublished predictions for direct orbits were in error by more than $30''$. The orbital elements published in ref. 21 and summarized in Table 1 use all available observations to the end of November 1997, and should accurately predict the positions of the moons in April 1998 to within a few arcminutes. These particular orbits were chosen with the intention of maximizing the eccentricity of U 1 and minimizing the eccentricity of U 2. Reliable orbits must await 1998 recovery observations. If the $\sim 150^\circ$ orbital inclinations prove correct, the eccentricities and inclinations will suffer secular Kozai oscillations²³ induced by solar perturbations, with amplitudes of 10–20% in eccentricity e and $\sin i$ (where i is inclination) on timescales of 2,000–5,000 yr.

Assuming geometric albedos of 0.07 (as measured for two of the small uranian ring moons²⁴), we estimate the radii of U 1 and U 2 at 30 and 60 km, respectively. The assumed albedo exceeds that of the jovian irregular satellites Himalia and Elara (~ 0.03) and Saturn's Phoebe (0.046–0.06) (ref. 1), but is much less than Nereid's albedo of ~ 0.20 (ref. 25); errors of a factor of 2 in radius are possible. The images taken on 6 and 7 September show no detectable brightness variations greater than ~ 0.05 mag for U 2 and ~ 0.10 mag for U 1, such as might be attributed to the rotation of an irregular body.

During the September observations we searched Neptune's vicinity using the identical technique. Although we detected Nereid ($m_V \approx 19$) at its predicted position, no other satellites of Neptune were found. Our upper limit (5σ) of $m_R = 23.5$ corresponds to a radius of ~ 35 km for a geometric albedo of 0.07.

Broadband photometry from our 27 October images (Table 1) indicates colours which are decidedly redder than the Sun (solar $(B - R) = 1.03$, solar $(R - I) = 0.33$). By contrast, the regular uranian satellites, rings and ring moons all show neutral or grey colours^{24,26}. Several irregular satellites of Jupiter and Saturn also have neutral spectra, similar to those of C-type asteroids^{27–29}. However, the recently discovered Centaur and Kuiper-belt objects show a wide range of reddish colours^{30,31} encompassing those of U 1 and U 2, and perhaps resulting from the bombardment of organic-rich icy surfaces by energetic particles^{32,33}. As the newly discovered moons were probably captured by Uranus soon after its formation⁷, their physical properties may provide valuable clues to conditions in the early Solar System. $\square$

Received 9 December 1997; accepted 11 February 1998.

1. Burns, J. A. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 1–38 (Univ. Arizona Press, Tucson, 1986).
2. Pollack, J. B., Lunine, J. & Titemore, W. C. in *Uranus* (eds Bergstrahl, J. T., Miner, E. D. & Matthews, M. S.) 469–512 (Univ. Arizona Press, Tucson, 1991).
3. Jewitt, D., Luu, J. & Chen, J. The Mauna Kea-Cerro Tololo Kuiper belt and Centaur survey. *Astron. J.* **112**, 1225–1238 (1996).
4. Kowal, C. T., Aksnes, K., Marsden, B. G. & Roemer, E. Thirteenth satellite of Jupiter. *Astron. J.* **80**, 460–464 (1975).
5. Pickering, W. H. *Ann. Harvard Coll. Obs.* **53**, 45–73, 85–142 (1905).
6. Kuiper, G. The second satellite of Neptune. *Publ. Astron. Soc. Pacif.* **61**, 175–176 (1949).
7. Pollack, J. B., Burns, J. A. & Tauber, M. E. Gas drag in primordial circumplanetary envelopes: A mechanism for satellite capture. *Icarus* **37**, 587–611 (1979).
8. Kuiper, G. in *Planets and Satellites* (eds Kuiper, G. & Middlehurst, B. M.) 575–591 (Univ. Chicago Press, Chicago, 1961).
9. Owen, W. M. Jr & Synnott, S. P. Orbits of the ten small satellites of Uranus. *Astron. J.* **93**, 1268–1271 (1987).
10. Christie, W. Report on a search for further satellites of the planets Uranus and Neptune. *Publ. Astron. Soc. Pacif.* **42**, 253 (1930).
11. Smith, B. in *Uranus and Neptune* (ed. Bergstrahl, J. T.) 213–223 (CCP-2330, NASA, Washington DC, 1984).
12. Cruikshank, D. P. & Brown, R. H. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 836–873 (Univ. Arizona Press, Tucson, 1986).
13. Żytkow, A., Irwin, M., Webster, R. & Tremaine, S. A photographic search for satellites of Uranus. *Icarus* **102**, 298–306 (1993).
14. Greenberg, R., Bottke, W. F., Carusi, A. & Valsecchi, G. B. Planetary accretion rates: Analytical derivation. *Icarus* **94**, 98–111 (1991).

15. Hamilton, D. P. & Krivov, A. V. Dynamics of distant moons of asteroids. *Icarus* **128**, 241–249 (1997).
16. Goldreich, P. Inclination of satellite orbits about an oblate precessing planet. *Astron. J.* **70**, 5–9 (1966).
17. Gladman, B. & Kavelaars, J. J. Kuiper belt searches from the Palomar 5-m telescope. *Astron. Astrophys.* **317**, 35–38 (1997).
18. Luu, J. Twice in a blue moon. *Nature* **390**, 441–443 (1997).
19. Gladman, B. et al. *IAU Circ.* No. 6764 (1997).
20. Fitzsimmons, A., Fletcher, M., Irwin, M., Offutt, W. & Hergenrother, C. *IAU Circ.* No. 6771 (1997).
21. Marsden, B. & Williams, G. V. *IAU Circ.* No. 6780 (1997).
22. Marsden, B. & Williams, G. V. *IAU Circ.* No. 6765 (1997).
23. Thomas, F. & Morbidelli, A. The Kozai resonance in the outer solar system and the dynamics of long-period comets. *Celest. Mech. Dyn. Astron.* **64**, 209–229 (1996).
24. Thomas, P. C., Weitz, C. & Veverka, J. Small satellites of Uranus: Disk-integrated photometry and estimated radii. *Icarus* **81**, 92–101 (1989).
25. Thomas, P. C., Veverka, J. & Helfenstein, P. Voyager observations of Nereid. *J. Geophys. Res.* **96**, 19253–19259 (1991).
26. Veverka, J., Brown, R. H. & Bell, J. F. in *Uranus* (eds Bergstrahl, J. T., Miner, E. D. & Matthews, M. S.) 528–560 (Univ. Arizona Press, Tucson, 1991).
27. Cruikshank, D. P. Near-infrared studies of the satellites of Saturn and Uranus. *Icarus* **41**, 246–253 (1980).
28. Tholen, D. J. & Zellner, B. Eight-color photometry of Hyperion, Iapetus and Phoebe. *Icarus* **53**, 341–347 (1983).
29. Tholen, D. J. & Zellner, B. Multicolor photometry of outer jovian satellites. *Icarus* **58**, 246–253 (1984).
30. Luu, J. & Jewitt, D. Color diversity among the Centaurs and Kuiper belt objects. *Astron. J.* **112**, 2310–2318 (1996).
31. Green, S. F. et al. Surface reflectance properties of distant solar system bodies. *Mon. Not. R. Astron. Soc.* **290**, 186–192 (1997).
32. Fink, U. et al. The steep red spectrum of 1992 AD: An asteroid covered with organic material? *Icarus* **97**, 145–149 (1992).
33. Wilson, P. D. *Models of Organic-Rich Surfaces in the Outer Solar System*. Thesis, Cornell Univ. (1997).

Acknowledgements. Observations at the Palomar Observatory were made as part of a continuing collaborative agreement between the California Institute of Technology and Cornell University. We thank the telescope operators and mountain staff at Palomar for their assistance; D. Tholen, A. Fitzsimmons, C. Hergenrother and S. Lilly for their confirming observations; and T. Yokoyama for numerical integrations of the provisional orbits. Support came from the Natural Sciences and Engineering Research Council of Canada (B.G. and J.J.K.), and NASA's Planetary Geology and Geophysics program (P.D.N. and J.A.B.).

Correspondence and requests for materials should be addressed to B.J.G. (e-mail: gladman@cita.utoronto.ca).

Accretion rate of cosmic spherules measured at the South Pole

Susan Taylor*, James H. Lever* & Ralph P. Harvey†

* US Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire 03755, USA

† Case Western Reserve University, Cleveland, Ohio 44106, USA

Micrometeorites are terrestrially collected, extraterrestrial particles smaller than about 1 mm, which account for most of the mass being accreted to the Earth^{1,2}. Compared with meteorites, micrometeorites more completely represent the Earth-crossing meteoroid complex^{3,4} and should include fragments of asteroids, comets, Mars and our Moon, as well as pre-solar and interstellar grains^{3,6}. Previous measurements of the flux of micrometeoroids that survive to the Earth's surface have large uncertainties owing to the destruction of particles by weathering^{7–9}, inefficiencies in magnetic collection or separation techniques^{7–9}, low particle counts^{10,11}, poor age constraint^{7–9,12,13} or highly variable concentrating processes^{12,13}. Here we describe an attempt to circumvent these problems through the collection of thousands of well preserved and dated micrometeorites from the bottom of the South Pole water well, which supplies drinking water for the Scott–Amundsen station. Using this collection, we have determined precise estimates of the flux and mass distribution for 50–700- μ m cosmic spherules (melted micrometeorites). Allowing for the expected abundance of unmelted micrometeorites¹⁴ in the samples, our results indicate that about 90% of the incoming mass of submillimetre particles evaporates during atmospheric entry. Our data indicate the loss of glass-rich and small stony spherules from deep-sea deposits^{7,8}, and they provide constraints for models describing the survival probability of micrometeoroids^{15,16}.

Precisely measuring the micrometeorite flux requires a large, unbiased deposit of particles, little or no weathering, precise age constraint, a known area of contribution, and an unbiased collection technique of known efficiency. Our collection of micrometeorites from the South Pole water well (SPWW) meets these requirements. The well is a 5,000-m³ subsurface water pool that supplies drinking water for the Scott–Amundsen station (Fig. 1). Micrometeorites originally on the snow surface became incorporated into essentially horizontal ice layers whose depth–age relationship has been determined from the stratigraphy of an ice core¹⁷. As the well melts downwards, micrometeorites released from the ice remain on the bottom of the water-filled cavity.

In December 1995, we collected micrometeorites from the SPWW using a robot that suctioned and filtered the bottom particle deposits (Fig. 1). The bottom of the SPWW was quite smooth (1-mm depressions over scales of 1–5 mm), but differential melting had formed a relatively flat central plateau surrounded by

sculpted pockets (steep dips and smaller plateaux). Most pockets contained dark concentrations of particulates; no particle concentrations were visible on the central plateau. One collector deployment suctioned the central plateau, and other deployments cleaned three adjacent pockets. Suctioning left the plateau and nearly all of the pocket areas visibly clean, indicating high collection efficiency¹⁸. The complex topography limited our collection area to 30 m².

We have processed the central-plateau sample and one pocket sample (Table 1). Most particles were iron oxide grains derived from a pump failure in the water-supply system. This contamination makes it difficult to quantify the abundance of unmelted micrometeorites, so our flux and size data are derived only from the easily separated cosmic spherules. Spherules were counted as extraterrestrial if they had major-element compositions and crystal textures seen in other cosmic spherule collections^{4,19–21}. The only spherical contaminants consisted of copper and plastic, and were easily separated. The samples show no signs of particle loss by dissolution:

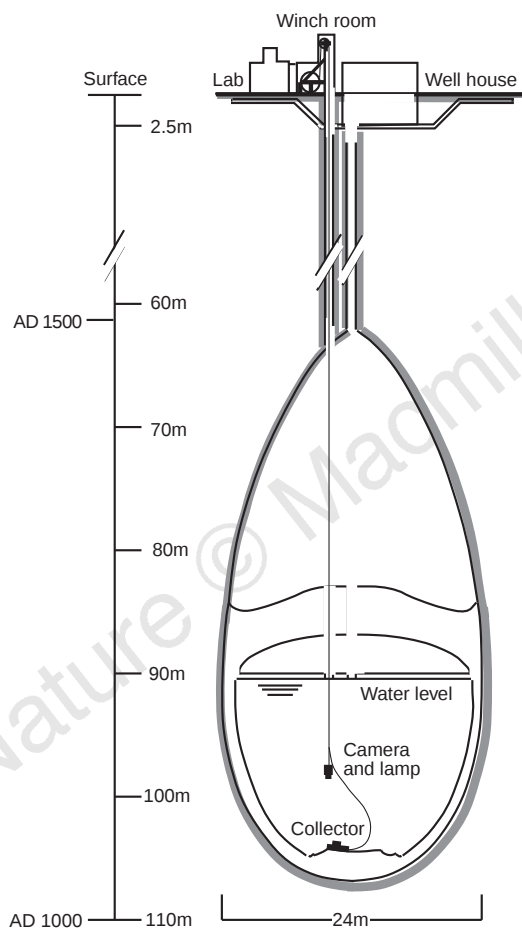


Figure 1 Approximate size and shape of the South Pole water well in December 1995. The well had melted ~8000 ton of firm and ice since its construction in December 1992. The age interval of the ice in years is $\Delta T = (10.3 \pm 0.1)\Delta d$, where d is the depth from 50 to 110 m (ref. 17). The central pump and hoses used to operate the well are not shown. We designed the collector to recover all micrometeorites in the size range 50–2,000 μm independent of their density, magnetic susceptibility or shape^{18,27}. It suctioned and internally filters the particles while traversing the well bottom. When tested using 70–400- μm quartz and steel particles, its collection efficiency exceeded 99% on smooth or moderately rough ice¹⁸. The collector developed intake flow velocities roughly 2.5 and 6 times higher, respectively, than the velocities needed to entrain 2,000- μm iron and stony spherules. We therefore expect that the high measured collection efficiency applies for all 50–2,000- μm micrometeorites. We used an underwater video camera to control and record the collector movement over the well bottom.

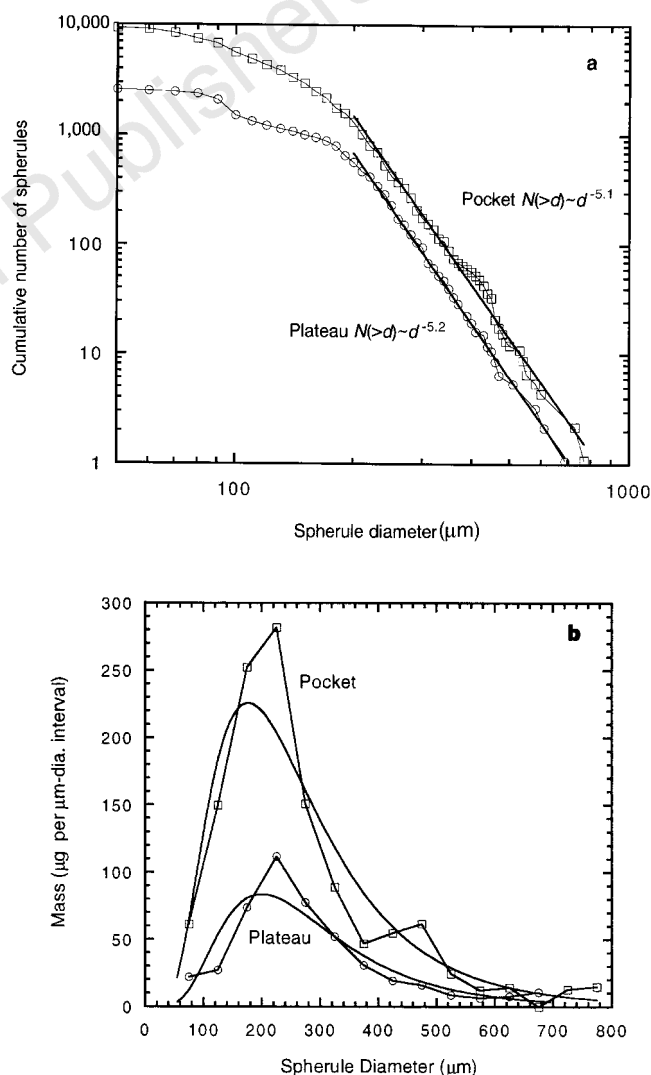


Figure 2 Size and mass distribution of SPWW spherules. **a**, Cumulative size distributions for the SPWW spherules based on 756 and 875 measured diameters for the plateau and pocket samples, respectively. **b**, Mass distributions for the plateau and pocket samples using 50- μm bin widths. The smooth curves are log-normal distributions fitted to the cumulative mass data and have the following parameters: plateau, $\bar{x} = 2.41$, $\sigma_x = 0.216$, $M_{50-700} = 0.0231$ g, $r = 0.995$; pocket, $\bar{x} = 2.37$, $\sigma_x = 0.233$, $M_{50-800} = 0.0613$ g, $r = 0.996$, where $x = \log_{10}d(\mu\text{m})$, $\bar{x}$ and σ_x denote mean and standard deviation, respectively, r is the correlation coefficient and M_{50-700} is the total mass over the size range 50–700 μm .

Table 1 Material examined for the SPWW flux calculations

	Sample			Sample		
	Plateau			Pocket		
Size fraction (μm)	53–106	106–250	>250	53–106	106–250	>250
Mass in size fraction (g)	2.93	5.13	8.85	12.7	23.8	12.4
Mass sorted (g)	0.248	3.50	8.85	0.349	3.62	12.4
Spherules counted	122	655	183	121	731	227
Expected count in size fraction	1,440	960	183	4,400	4,810	227

Material from the collector filters was back-flushed with well water into stainless-steel sieves, yielding three size fractions per sample. We sorted and extracted cosmic spherules from all of the material in the >250- μm size fractions. However, because the smaller size fractions are more difficult to examine, we sorted only a portion of each. The expected spherule count assumes that spherules occur throughout a size fraction in the same concentration as the mass sorted. We estimate that errors resulting from concentration variations, particle loss during sorting, and misidentification of spherules combine to yield <10% uncertainty in the expected spherule counts.

compared with collections from the deep sea^{3,8} and Greenland^{13,22}, easily weathered pure glass spherules are more abundant (20% versus 1–10%) and stony spherules show much less etching of interstitial glass. Also, the presence of fragile, bubble-rich glass particles and the absence of spherule fragments indicate that few particles were broken during collection.

Using an optical microscope, we measured particle major and minor axes and averaged these to determine spherule diameters. Most particles were spherical (84% of the particles had major and minor axes equal to within $\pm 20\%$). Figure 2a shows the resulting cumulative size distributions. For diameters larger than 200 μm , the slopes of these distributions are -5.2 ± 0.1 for the plateau sample and -5.1 ± 0.2 for the pocket sample. These slopes are steeper than the -3.83 determined by Murrell *et al.*⁷ for 200–700- μm deep-sea stony spherules, consistent with destruction of smaller spherules by prolonged exposure to sea water.

The size distributions were converted to mass distributions (Fig. 2b) assuming an average particle density of $2.8 \pm 0.3 \text{ g cm}^{-3}$ (estimated from the proportion of each spherule type and their reported⁷ or estimated densities). Image processing of video records taken during collector deployment yielded the areas suctioned ($\pm 5\%$) as 17 m^2 for the plateau sample and 4.7 m^2 for the pocket sample. The age interval of the ice that contributed particles was $430 \pm 50 \text{ yr}$, based on measured age–depth data¹⁷ and allowing uncertainty for the depth at which the particle deposits first form. The two samples yield significantly different spherule fluxes: $(3.2 \pm 0.6) \times 10^{-6} \text{ g m}^{-2} \text{ yr}^{-1}$ for the central plateau and $(30 \pm 6) \times 10^{-6} \text{ g m}^{-2} \text{ yr}^{-1}$ for the pocket.

Evidence that the central plateau preserves the original surface flux of 50–2,000- μm micrometeorites derives from data obtained during collector deployment¹⁸ and results of computational fluid dynamics (CFD) modelling. First, the well circulation (Fig. 3) cannot move micrometeorites off or onto the central plateau: (i) the collector recovered submicrometre iron oxide grains (fall velocities $< 2 \mu\text{m s}^{-1}$) from the plateau and pocket, and these would only settle and remain on the well bottom for conditions of slow, laminar flow, (ii) CFD-model surface shear stresses are at least an order of magnitude smaller than the threshold needed to transport 50- μm micrometeorites along a horizontal surface²³, (iii) the CFD model shows that particles released along the well sides are not entrained but roll downward and collect in the bottom pockets, and (iv) extraterrestrial material was not found in water pumped from the middle of a similar polar well^{24,25}, probably because the micrometeorites were not entrained and remained on the well bottom. Second, no pocket remnants existed on the plateau at the time of collection, and because flow cannot transport particles off the plateau, we infer that earlier pockets did not overlie the plateau. That is, the central plateau contained micrometeorites only from the ice vertically above it. Third, $\sim 75\%$ of the projected area of the well bottom has a slope steeper than 45° (ref. 18), adequate to account for the order-of-magnitude enrichment of the pocket sample. The large spherules in the pocket sample and the lack of size sorting (Fig. 2a) also favour rolling of spherules down the sides over transport off the plateau.

In Fig. 4 we compare the plateau spherule flux to the micrometeorite flux measured in Greenland¹² and the micrometeoroid influx measured using the Long Duration Exposure Facility (LDEF) satellite¹. The Greenland spherule flux agrees with the SPWW flux above $\sim 150 \mu\text{m}$. The discrepancy for smaller spherules could be due to size sorting by flowing water that created some of the Greenland deposits. Because of the evidence already presented, we think that size sorting does not occur in the SPWW. Also, the low SPWW spherule flux below $\sim 100 \mu\text{m}$ is consistent with indications that most micrometeorites smaller than 100 μm survive atmospheric entry as unmelted particles^{14,15}.

Atmospheric heating effects make it more difficult to compare the

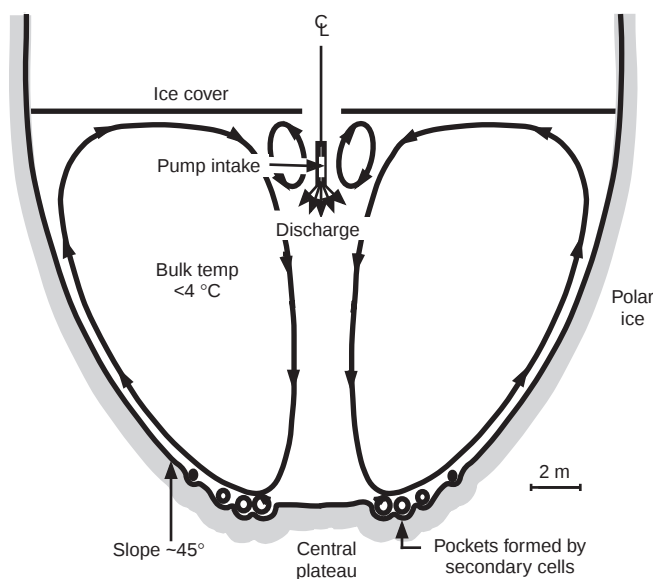


Figure 3 Circulation pattern in the SPWW. Although no velocity data exist, the pattern shown is consistent with other data and CFD analyses: (1) Low water injection rates (1 l s^{-1} at 20°C discharged in a 90° conical pattern 13 m above the well bottom) ensure that natural convection, driven by heat transfer along the ice, governs the flow field. (2) A measured centre-line temperature of 2.4°C and an ice cover on the pool are consistent with a descending ‘warm’ plume, and upward circulation along the walls as the water cools and becomes less dense. (3) The sharp-edged pockets and particulate patches along the inner, rising edges of the pockets indicate that they are sculpted by counter-rotating cells driven by the main circulation pattern. (4) An axisymmetric CFD model of the SPWW confirms the natural-convection pattern shown here (J.H.L., Richter, H., Kolodny, J. and Rigney, M., manuscript in preparation). The CFD model used the commercially available software Fluent²⁸ and included the density–temperature relationship for water near its maximum density ($\sim 4^\circ\text{C}$) to provide buoyancy forcing. Verification runs duplicated published natural convection flow fields for water near 4°C (refs 29, 30). With SPWW geometry, inlet flow conditions and 0°C wall temperatures specified, the software numerically solved the Navier–Stokes equations of motion until steady-state conditions were achieved. The model flow field was laminar, and relatively insensitive to the inflow velocity, temperature and pattern.

SPWW flux and the LDEF influx. SPWW spherules were probably 1.5–2 times larger in diameter when they entered the atmosphere¹⁵, so spherules larger than ~250 μm were derived from particles larger than those detected by the LDEF¹. Also, low-velocity, shallow-entry-angle micrometeoroids preferentially survive¹⁵, yet these particles contribute least to the total mass influx. Instead, the SPWW spherule flux and size distribution constrain models that describe atmospheric heating of micrometeoroids^{15,16}, so that the initial population creates the spherule population measured here. For example, Love and Brownlee¹⁵ predict that evaporative losses during atmospheric entry steepen the slope of the micrometeoroid size distribution by a power of -1.0 for particles larger than 100 μm . If this heating model is correct, the slope of -5.2 ± 0.1 measured for the plateau sample (Fig. 2a) could develop from the micrometeoroid size distribution obtained by Grün *et al.*²⁶ (slope of -4.0) but not from distribution derived from the LDEF data¹ (slope of -5.8).

Integration of the LDEF influx distribution (ref. 1, Fig. 4) yields an accretion rate for 20–400- μm micrometeoroids of $30,000 \text{ ton yr}^{-1}$. Love and Brownlee¹ reported $40,000 \pm 20,000 \text{ ton yr}^{-1}$ to account for all cosmic dust (S. G. Love, personal communication). This value agrees with the average extraterrestrial accretion rate over the past 80 million years derived from marine osmium², $37,000 \pm 13,000 \text{ ton yr}^{-1}$. We calculate, on the basis of the plateau sample and assuming a uniform global distribution, a terrestrial accretion rate for 50–700- μm cosmic spherules of $1,600 \pm 300 \text{ ton yr}^{-1}$, or $4 \pm 2\%$ of the micrometeoroid accretion rate. Our rate is higher than the rates estimated by Kyte⁸ from deep-sea stony spherules (740 ton yr^{-1}) and the deep-sea iron spherules of Murrell *et al.*⁷ corrected for post-depositional weathering of stony spherules ($400\text{--}1,200 \text{ ton yr}^{-1}$)⁸. Dissolution of the glass and cryptocrystalline spherules in the proportions found in the SPWW samples (33%) would account for these differences. Our result overlaps estimates by Yiou and Raisbeck¹⁰ and Yiou *et al.*¹¹ based on 5 and 14 spherules, respectively, extracted from well-dated Antarctic ice cores ($1,000\text{--}1,500 \text{ ton yr}^{-1}$), and it falls within the broad range calculated by Peng and Lui⁹ based on spherules extracted from deep-

sea sediments ($500\text{--}7,000 \text{ ton yr}^{-1}$). The SPWW spherule accretion rate in the size range 50–300 μm is $1,100 \pm 200 \text{ ton yr}^{-1}$, lower than the $2,200 \text{ ton yr}^{-1}$ measured in Greenland¹².

Although we were unable to quantify the number of unmelted micrometeorites in the SPWW samples, the collections of Maurette *et al.*^{12,14} allow us to estimate their abundance. The Greenland collection contained ~25–30% unmelted micrometeorites in each size fraction¹². However, unmelted micrometeorites obtained from Antarctic blue ice constituted ~40% of the extraterrestrial mass collected, and these fell mostly in the 50–100 μm size fraction, where they outnumbered spherules by a factor of five (ref. 14). Lower terrestrial contamination in this size fraction in Antarctica may explain the higher ratio¹³. If the SPWW samples contain unmelted micrometeorites in the proportion found in Antarctic blue ice¹⁴, our accretion rate would increase to $2,700 \pm 1,400 \text{ ton yr}^{-1}$. Thus, ~90% of the micrometeoroid influx evaporates during atmospheric entry.

The SPWW is a closed system that contains all the meteoritic material that fell on the pre-industrial snow now being melted. Its large volume and good stratigraphic control allow precise measurement of the flux and size distribution of micrometeorites. The particles themselves represent an enormous, well-preserved and unbiased sample of the materials accreting to Earth. As such, they provide a standard for determining the effects of atmospheric heating and terrestrial weathering of submillimetre extraterrestrial materials. Furthermore, the SPWW progressively melts into older ice; annual collections will assess flux and composition variations on 100-yr intervals to about AD 400. Serial collections will also clean the well of iron oxide contamination, making it feasible to sort and quantify unmelted micrometeorites. □

Received 10 September 1997; accepted 16 March 1998.

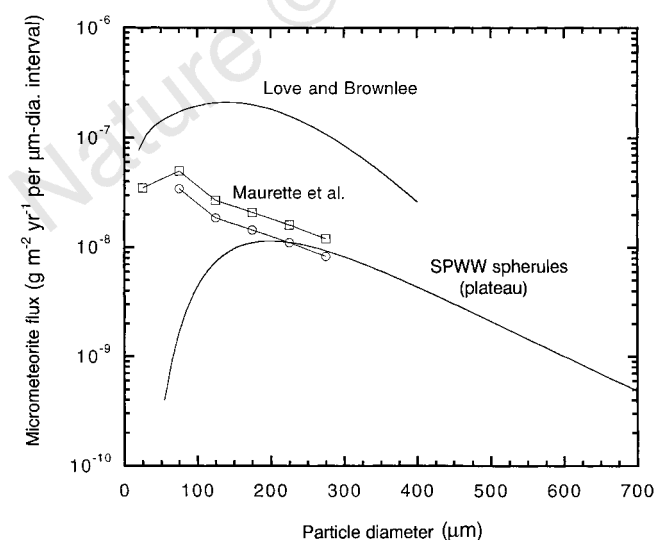


Figure 4 SPWW spherule flux compared with the flux measured above the atmosphere¹ and in Greenland¹². About $\pm 20\%$ uncertainty applies to the SPWW data, and about $\pm 50\%$ applies to the data of Love and Brownlee¹. While Maurette *et al.*¹² give no value, the uncertainty in their data must be large because of the poorly known age and variable concentration of the Greenland deposits. The circles represent the spherule flux measured in Greenland, while the squares show the total micrometeorite flux. Assuming a uniform global distribution, the SPWW flux implies a terrestrial accretion rate for 50–700- μm cosmic spherules of $1,600 \pm 300 \text{ ton yr}^{-1}$.

1. Love, S. G. & Brownlee, D. E. A direct measurement of the terrestrial mass accretion rate of cosmic dust. *Science* **262**, 550–553 (1993).
2. Peucker-Ehrenbrink, B. Accretion of extraterrestrial matter during the last 80 million years and its effect on the marine osmium isotope record. *Geochim. Cosmochim. Acta* **60**, 3187–3196 (1996).
3. Brownlee, D. E. in *The Sea* (ed. Emiliani, C.) Ch. 19 (Wiley, New York, 1981).
4. Brownlee, D. E., Bates, B. & Schramm, L. The elemental composition of stony cosmic spherules. *Meteorit. Planet. Sci.* **32**, 157–175 (1997).
5. Bradley, J. P., Sandford, S. A. & Walker, R. M. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 861–895 (Univ. Arizona Press, Tucson, 1988).
6. Brownlee, D. E. *et al.* Identification of cometary and asteroidal particles in stratospheric IDP collections. *Lunar Planet. Sci.* **XXIV**, 205–206 (1993).
7. Murrell, M. T., Davis, P. A., Nishizumi, K. & Millard, H. T. Deep-sea spherules from Pacific clay: Mass distribution and influx rate. *Geochim. Cosmochim. Acta* **44**, 2067–2074 (1980).
8. Kyte, F. T. *Analyses of Extraterrestrial Materials in Terrestrial Sediments* Thesis (Univ. California, Los Angeles (1983).
9. Peng, H. & Lui, Z. Measurement of the annual flux of cosmic dust in deep-sea sediments. *Meteoritics* **24**, 315 (1989).
10. Yiou, F. & Raisbeck, G. M. Cosmic spherules from an Antarctic ice core. *Meteoritics* **22**, 539–540 (1987).
11. Yiou, F., Raisbeck, G. M. & Jehanno, C. Influx of cosmic spherules to the Earth during the last ~10⁵ years as deduced from concentrations in Antarctic ice cores. *Meteoritics* **24**, 344 (1989).
12. Maurette, M., Jehanno, C., Robin, E. & Hammer, C. Characteristics and mass distribution of extraterrestrial dust from the Greenland ice cap. *Nature* **328**, 699–702 (1987).
13. Maurette, M., Hammer, C. & Pourchet, M. in *From Mantle to Meteorites* (eds Gopalan, K., Gaur, V. K., Somayajulu, B. L. & MacDougall, F. D.) 87–126 (Indian Acad. Sciences, Bangalore, 1990).
14. Maurette, M. *et al.* A collection of diverse micrometeorites recovered from 100 tonnes of Antarctic blue ice. *Nature* **351**, 44–47 (1991).
15. Love, S. G. & Brownlee, D. E. Heating and thermal transformation of micrometeoroids entering the earth's atmosphere. *Icarus* **89**, 26–43 (1991).
16. Flynn, G. J. Atmospheric entry heating: A criterion to distinguish between asteroidal and cometary sources of interplanetary dust. *Icarus* **77**, 287–310 (1989).
17. Kuivinen, K. C., Koci, B. R., Holdsworth, G. W. & Gow, A. J. South Pole ice core drilling, 1981–1982. *Antarct. J. US XVII*, 89–91 (1982).
18. Taylor, S., Lever, J. H., Harvey, R. P. & Govoni, J. *Collecting Micrometeorites from the South Pole Water Well* (Rep. 97-1, Cold Regions Res. and Eng. Lab., Hanover, New Hampshire, 1997).
19. Blanchard, M. B., Brownlee, D. E., Bunch, T. E., Hodge, P. W. & Kyte, F. T. Meteoroid ablation spherules from deep sea sediments. *Earth Planet. Sci. Lett.* **46**, 178–190 (1980).
20. Koeberl, C. & Hagen, E. H. Extraterrestrial spherules in glacial sediment from the Transantarctic Mountains, Antarctica: Structure, mineralogy and chemical composition. *Geochim. Cosmochim. Acta* **53**, 937–944 (1989).
21. Harvey, R. P. & Maurette, M. The origin and significance of cosmic dust from the Walcott Neve, Antarctica. *Proc. Lunar Planet. Sci. Conf.* **21**, 569–578 (1991).
22. Maurette, M., Hammer, C., Brownlee, D. E., Reeh, N. & Thomsen, H. H. Placers of cosmic dust in the blue ice lakes of Greenland. *Science* **233**, 869–872 (1986).
23. *Sedimentation Engineering* (ed. Vanoni, V. A.) 91–96 (ASCE Manuals and Reports on Engineering Practice No. 54, ASCE, New York, 1975).
24. McCorkell, R. H., Pinson, W. H., Fireman, E. L. & Langway, C. C. Jr A search for cosmic dust in a large collection of particulate and dissolved material from polar ice. *Int. Assoc. Sci. Hydrol.* **86**, 25–30 (1970).

25. Fireman, E. L. & Langway, C. C. Jr Search for aluminum-26 in dust from the Greenland ice sheet. *Geochim. Cosmochim. Acta* **29**, 21–27 (1965).
26. Grun, E., Zook, H. A., Fechtig, H. & Geise, R. H. Collisional balance of the meteoritic complex. *Icarus* **62**, 244–272 (1985).
27. Lever, J. H., Taylor, S. & Harvey, R. P. A collector to retrieve micrometeorites from the South Pole water well. *Lunar Planet. Sci. XXVII*, 747–748 (1996).
28. Fluent Version 4.3, Fluent, Inc., Lebanon, New Hampshire 03766, USA (1995).
29. Lin, D. S. & Nansteel, M. W. Natural convection in a vertical annulus containing water near the density maximum. *J. Heat Transfer* **109**, 899–905 (1987).
30. Lankford, K. E. & Bejan, A. Natural convection in a vertical enclosure filled with water near 4 °C. *J. Heat Transfer* **108**, 755–763 (1986).

Acknowledgements. We thank J. Rand, J. Govoni and M. Shandrick for their help with our field work, and C. Engrand, D. Joswiak, S. Kuehner, D. Brownlee, M. Maurette, G. Kurat and C. Daghljan for their help with particle analysis. This work was supported by the National Science Foundation with additional support from CRREL.

Correspondence and requests for materials should be addressed to J.H.L. (e-mail: jlever@crrel.usace.army.mil).

Charge separation in localized and delocalized electronic states in polymeric semiconductors

A. Köhler*, D. A. dos Santos†, D. Beljonne‡, Z. Shuai†, J.-L. Brédas†, A. B. Holmes‡, A. Kraus§, K. Müllen§ & R. H. Friend*

* Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

† Service de Chimie des Matériaux Nouveaux, Centre de Recherche en Electronique et Photonique Moléculaires, Université de Mons-Hainaut, Place du Parc, 20, B-7000 Mons, Belgium

‡ University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK

§ Max-Planck-Institut für Polymerforschung, Ackermannweg 10, 55128 Mainz, Germany

Conjugated polymers such as poly(*p*-phenylene vinylene)s (PPVs) allow low-cost fabrication of thin semiconducting films by solution processing onto substrates. Several polymeric optoelectronic devices have been developed in recent years, including field-effect transistors¹, light-emitting diodes², photocells^{3,4} and lasers⁵. It is still not clear, however, whether the description of electronic excitations in these materials is most appropriately formulated within a molecular or semiconductor (band-theory) picture. In the former case, excited states are localized and are described as excitons; in the latter they are delocalized and described as free electron–hole pairs. Here we report studies of the electronic states associated with optical excitations in the visible and ultraviolet range for the conjugated polymer poly(2-methoxy-5-(2'-ethyl-hexyloxy)-*p*-phenylene vinylene) (MEH-PPV), by means of photocurrent measurements and quantum-chemical calculations. We find several photocurrent spectral features between 3 and 5 eV which are coupled with bands in the absorption spectrum. On modelling the excited states in this energy range, we have discovered an important feature that is likely to be general for materials composed of coupled molecular units: that mixing of

delocalized conduction- and valence-band states with states localized on the molecular units produces a sequence of excited states in which positive and negative charges can be separated further at higher energies. In other words, these excited states facilitate charge separation, and provide a conceptual bridge between the molecular (localized) and semiconductor (delocalized) pictures.

An indication of the difficulty in finding a coherent description for polymeric materials which show both molecular and semiconductor properties is provided by the conflicting descriptions of polymer photoexcitations in the literature, ranging from 'free-electron' models⁶ to localized excitons^{7–9}. The question of how to describe such photoexcitations is not of merely academic interest, but determines the possible efficiency of polymer light-emitting diodes², diode lasers and photocells³ and can provide clues for the understanding of long-distance (>10 Å) electron-transfer processes that are discussed as possible within a DNA matrix and might play a role in the mechanisms of DNA damage and repair processes^{10,11}.

Our present photocurrent measurements on MEH-PPV (Fig. 1)

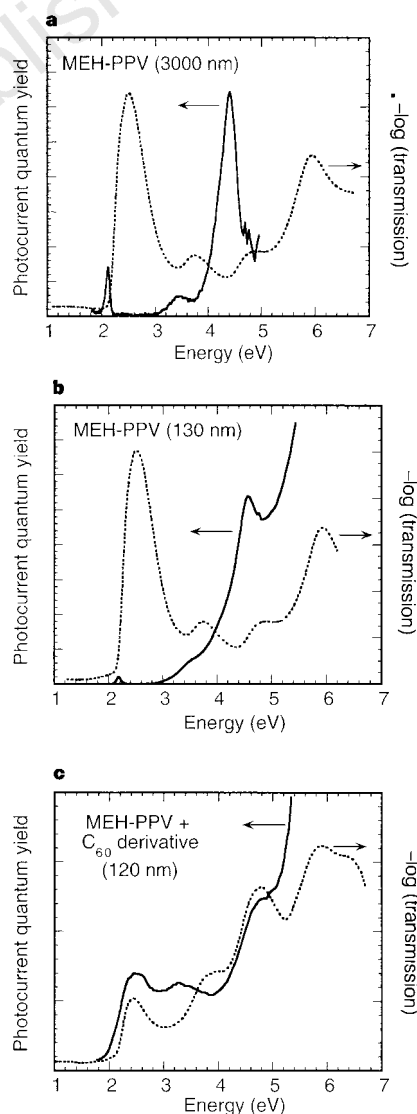


Figure 2 Short-circuit photocurrent spectra (solid line) and absorption spectra (dotted line) of different films. **a**, MEH-PPV (film thickness 3,000 nm); **b**, MEH-PPV (film thickness 130 nm); and **c**, a blend of 50 wt% MEH-PPV and the C₆₀ derivative. For **c**, an indium tin oxide (ITO) electrode was used instead of gold, and illumination was through the semitransparent aluminium electrode. Intensities are given in arbitrary units.

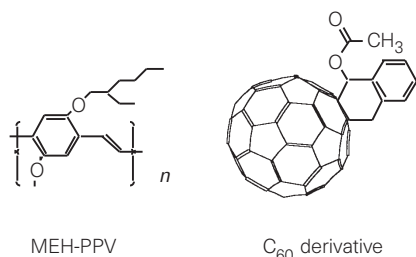


Figure 1 Structural formula of MEH-PPV and of the C₆₀ derivative.

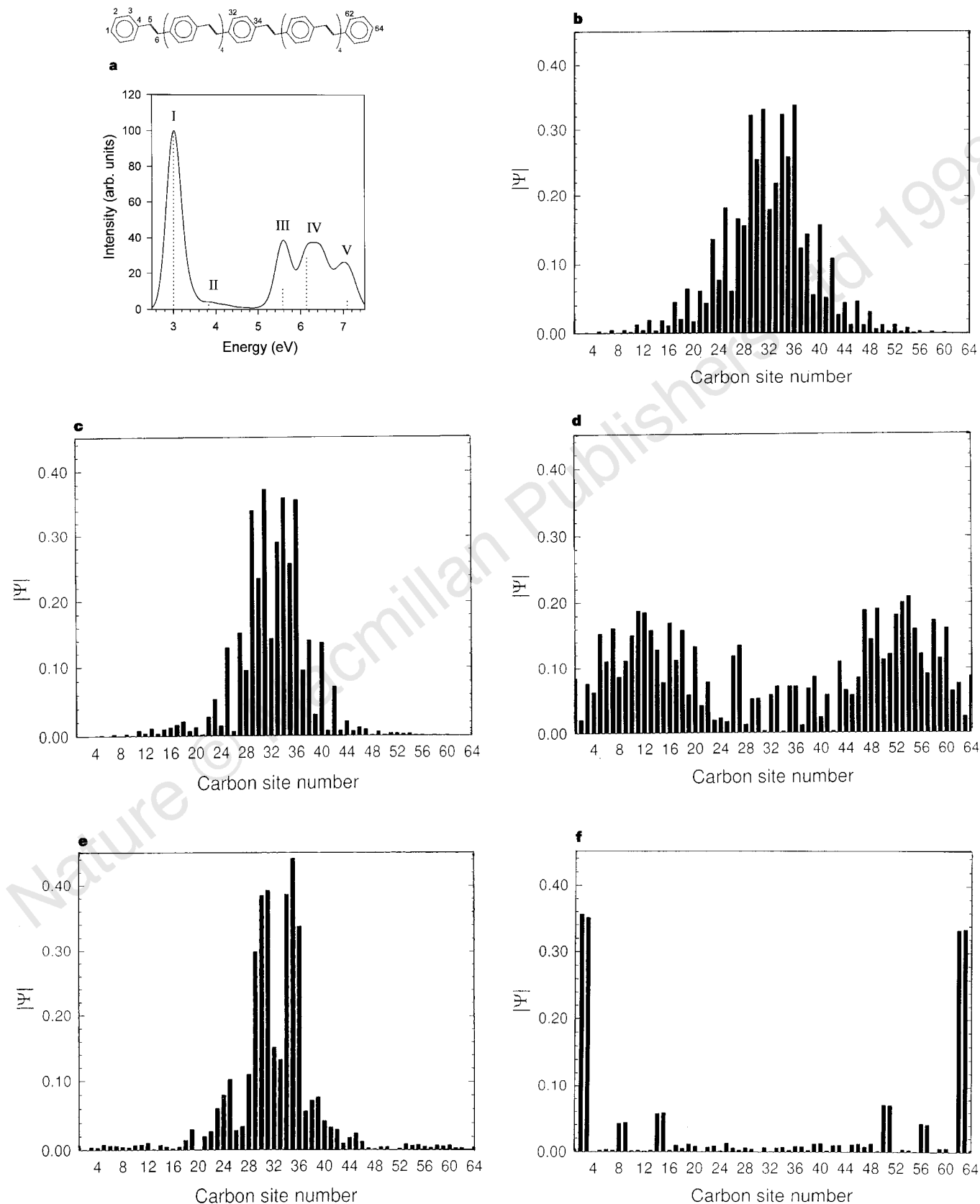


Figure 3 INDO/SCI calculations for the 11-ring PPV model oligomer; the site labelling is shown at the top. **a**, Simulated linear absorption spectrum. The vertical dotted lines represent the oscillator strength of the excited states for which the wavefunctions have been computed; the absolute value of the exciton wavefunction $|\Psi(x_e, x_h = 34)|$ calculated for the PPV undecamer as a function of carbon site for a fixed position of the hole at site 34 is shown for excitation into; **b**,

the first absorption peak (3.1 eV); **c**, the second absorption peak (3.9 eV); **d**, the third absorption peak (4.8 eV); **e**, the fourth absorption peak (6.2 eV); and **f**, the fifth absorption peak (>6.5 eV). The energies given refer to the experimentally found peak positions. The corresponding theoretical transition energies (eV) are at 3.0, 3.8, 5.6, 6.3 and 7.0.

extend earlier reports^{12,13} of the action spectrum for photodiodes and photoconduction spectra further into the ultraviolet, and allow comparison with the optical absorption spectra beyond the first π – π^* absorption band to cover three higher-lying absorption bands. The devices used for the photocurrent measurements were in the form of ‘sandwich’ structure diodes, fabricated in a layer structure of quartz substrate/semitransparent gold electrode/MEH-PPV film/semitransparent aluminium electrode. The short-circuit photocurrent was recorded in vacuum of 10^{-6} mbar. The photocurrent is corrected for the incident photon flux. Illumination was through the quartz substrate.

Figure 2a shows the photocurrent and absorption spectra of a photocell with a 3,000-nm-thick MEH-PPV film. The photocell was annealed at 100 °C for 12 h before the measurement to reduce the content of residual oxygen and water. We find photocurrent features at the onset of each absorption peak. In many photocells, we also observe a general photocurrent rise at and beyond 5 eV (Fig. 2b, for a 130-nm-thick MEH-PPV film, not annealed). We note that the third photocurrent feature is much larger than the first two. This increase in the photocurrent quantum yield does not scale with the size of the absorption coefficient, and cannot be accounted for by ‘internal filter effects’^{14–16}. For example, the absorption coefficient of MEH-PPV (Fig. 2b) is the same at 2.3 eV, 3.5 eV and 4.6 eV. Addition of an electron-accepting species such as a soluble C_{60} derivative¹⁷ (to give a blend of 50 wt%, corresponding to one C_{60} derivative per four MEH-PPV repeat units) increases the intensity of the first two photocurrent features relative to the third feature in addition to giving an overall rise in the photocurrent quantum yield (Fig. 2c). The rise in quantum yield will also involve other, complex contributions not addressed here, including improved charge transport. The photocurrent quantum yields were $\sim 10^{-3}\%$ at the first photocurrent peak for MEH-PPV; a value of 7% was measured for the blend of MEH-PPV and the C_{60} derivative when illuminated through the indium tin oxide electrode at a photon energy of 2.5 eV. The feature common to all these devices is the presence of several distinct photocurrent features connected with the absorption peaks. This demonstrates the role of several distinct electronic transitions in the process of charge generation, with different efficiencies of charge separation.

There has been considerable work done to calculate the optical transitions in PPV and related materials, by including Coulomb interactions¹³, and taking into account the full chemical structure of the polymer repeat unit^{18,19}. To investigate the cause of the large photocurrent yields at high photon energies we have performed ‘intermediate neglect of differential overlap’ (INDO)/‘single configuration interaction’ (SCI) calculations of the exciton wavefunction for the 11-ring PPV oligomer. We have computed the exciton wavefunction $|\Psi_{PPV}(x_e, x_h = 34)|$, which represents the probability amplitude of finding an electron in a given site x_e assuming that the hole is located in the middle of the oligomer backbone, that is, on site $x_h = 34$ (see the site labelling in Fig. 3a). We note that all occupied and unoccupied π levels are taken into account, in contrast to our earlier calculations¹⁸. Figure 3a shows the INDO/SCI simulated linear absorption spectrum of the model oligomer; representative exciton wavefunctions for each of the five calculated absorption peaks (hereafter denoted I, II, III, IV and V) are shown in Fig. 3b–f. These absorption peaks arise from contributions of both delocalized and localized π wavefunctions: the wavefunctions are illustrated in Fig. 4. When the electronic wavefunction of the molecular orbital has nodes at the *para*-positions of the ring, that is, at those carbon atoms which link neighbouring rings, there is negligible delocalization of charges between neighbouring rings—the charges are confined or localized to their own ring. However, when the electronic wavefunction has its nodes orthogonal to the polymer axis, the carbon atoms at the *para*-position carry a high charge density. Charge can then delocalize between neighbouring rings and indeed over the whole chain.

From Fig. 3b–f, we can distinguish between two types of excited states. First, there are states where the wavefunction, $|\Psi_{PPV}(x_e, x_h = 34)|$, has a narrow gaussian shape, centred on the hole site and confined to the neighbouring repeat units. This is the case for the states created by excitation into first, second and fourth absorption peaks. In particular, we estimate the coherence length of the lowest excited species (peak I), responsible for the luminescent properties of the polymer, to be of the order of five phenyl rings, while peak IV corresponds to excitations that hardly extend over more than a monomer unit. We find that the electronic transitions to these states are polarized parallel to the chain, consistent with the results reported in refs 19 and 20. From the analysis of the configuration interaction expansion, we find that these excited states arise from excitations between the same types of orbitals, that is, from delocalized occupied orbitals to delocalized unoccupied orbitals (peaks I and II), or between occupied and unoccupied electronic levels that are both confined on the phenylene rings (peak IV). We note that the excited states belonging to peaks I and IV correspond, respectively, to the ‘B’ and ‘D’ transitions mentioned in ref. 19. These are the excited states for which we measure only a low photocurrent quantum yield.

Second, there are excited states, which we term charge-transfer states, where there is a finite probability of finding the electron and hole separated by a few phenylene rings. This is the case for the states formed by excitation into the third (III) and fifth (V) absorption peak. These states correspond to excitations polarized in the plane of the π -network, with strong contributions perpendicular to the chain axis, as also reported in refs 19 and 20. In contrast to the first kind of states, their configuration-interaction description involves only electronic transitions from delocalized occupied molecular orbitals to localized unoccupied levels and vice versa (the electron is in a delocalized band state and the hole is confined on a benzene ring or vice versa; excitations ‘A’ and ‘C’ in ref. 19). These are the excited states that give rise to a particularly large photocurrent.

Sub-picosecond time-resolved measurements of photoluminescence have shown that there is extremely rapid cooling of higher-energy excited states, so that emission is essentially from the lowest singlet excited state within 1 ps (refs 21, 22). Under these conditions, the probability of charge separation depends on the ability of electron and hole to separate at short times (possibly aided by extrinsic means^{23–25}), before the lowest-lying, bound exciton is formed. We consider that this is facilitated in the higher-lying excitonic transitions when there is increased separation of electron and hole; this is particularly important for the third absorption band, as shown in Fig. 3. By addition of a high concentration of a highly electronegative compound, such as 50 wt% of a C_{60} derivative, efficient charge separation can also be achieved for peaks I and II. The increase in charge separation efficiency for excitation into

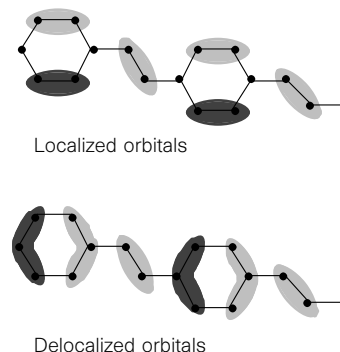


Figure 4 Scheme of molecular orbitals in poly(*p*-phenylene vinylene) with nodes at *para*-position and with nodes orthogonal to the molecular axis (after ref. 28). Different grey levels denote positive and negative signs of the molecular wavefunction.

peak I can be used for the fabrication of efficient polymer photo-voltaic cells⁴.

In excited states that have their origin in a transition from a delocalized orbital to a delocalized orbital, both carriers, electron and hole, are similarly extended, and therefore spatially overlapped. The same is true for excited states that involve transitions from a localized orbital to another localized orbital. However, in an excited state formed by a transition from a localized orbital to a delocalized orbital, the hole is confined to the phenylene ring, while the electron is delocalized over the chain (or vice versa). We consider that this is the reason for the higher photocurrent yields of these states.

These results, which have antecedents in early work on molecular crystals (refs 25–27 and refs therein), have more general implications for the understanding of electronic excitations in low-dimensional systems^{10,11}. The presence of charge-transfer states that separate more readily into free electron and holes can provide a natural description for the progression from molecular to semiconductor models of electronic excitations. □

Received 27 August 1997; accepted 23 February 1998.

- Garnier, F., Hajlaoui, R., Yassar, A. & Srivastava, P. All-polymer field-effect transistor realized by printing techniques. *Science* **265**, 1684–1686 (1994).
- Burroughes, J. H. *et al.* Light-emitting diodes based on conjugated polymers. *Nature* **347**, 539–541 (1990).
- Halls, J. J. M. *et al.* Efficient photodiodes from interpenetrating polymer networks. *Nature* **376**, 498–500 (1995).
- Yu, G., Gao, J., Hummelen, J. C., Wudl, F. & Heeger, A. J. Polymer photovoltaic cells: Enhanced efficiencies via a network of internal donor–acceptor heterojunctions. *Science* **270**, 1789–1791 (1995).
- Tessler, N., Denton, G. J. & Friend, R. H. Lasing from conjugated-polymer microcavities. *Nature* **382**, 695–697 (1996).
- Lee, C. H., Yu, G., Moses, D. & Heeger, A. J. Picosecond transient photoconductivity in poly(p-phenylenevinylene). *Phys. Rev. B* **49**, 2396–2407 (1994).
- Heun, S. *et al.* Conformational effects in poly(p-phenylene vinylene) revealed by low-temperature site-selective fluorescence. *J. Phys.: Condens. Matter* **5**, 247–260 (1993).
- Rauscher, U., Bässler, H., Bradley, D. D. C. & Hennecke, M. Exciton versus band description of the absorption and luminescence spectra in poly(para-phenylenevinylene). *Phys. Rev. B* **42**, 9830–9836 (1990).
- Brédas, J. L., Cornil, J. & Heeger, A. J. The exciton binding energy in luminescent conjugated polymers. *Adv. Mater.* **8**, 447–452 (1996).
- Frommherz, P. & Rieger, B. Photoinduced electron transfer in DNA matrix from intercalated ethidium to condensed methylviologen. *J. Am. Chem. Soc.* **108**, 5361–5362 (1986).
- Hall, D. B., Holmlin, R. E. & Barton, J. K. Oxidative DNA damage through long-range electron transfer. *Nature* **382**, 731–735 (1996).
- Pichler, K. *et al.* Optical spectroscopy of highly ordered poly(p-phenylene vinylene). *J. Phys.: Condens. Matter* **5**, 7155–7172 (1993).
- Chandross, M. *et al.* Excitons in poly(para-phenylenevinylene). *Phys. Rev. B* **50**, 14702–14705 (1994).
- Tang, C. W. & Albrecht, A. C. Photovoltaic effects of metal-chlorophyll-metal sandwich cells. *J. Chem. Phys.* **62**, 2139–2149 (1974).
- Marks, R. N., Halls, J. J. M., Bradley, D. D. C., Friend, R. H. & Holmes, A. B. The photovoltaic response in poly(p-phenylene vinylene) thin-film devices. *J. Phys.: Condens. Matter* **6**, 1379–1394 (1994).
- Harrison, M. G., Grüner, J. & Spencer, G. C. W. Analysis of the photocurrent action spectra of MEH-PPV polymer photodiodes. *Phys. Rev. B* **55**, 7831–7849 (1997).
- Kraus, A., Gügel, A., Belik, P., Walter, M. & Müllen, K. Covalent attachment of various substituents in close proximity of the C60-core: A broad synthetic approach to stable fullerene derivatives. *Tetrahedron* **51**, 9927–9940 (1996).
- Cornil, J., Beljonne, D., Friend, R. H. & Brédas, J. L. Optical absorptions in poly(para-phenylene vinylene) and poly(2,5-dimethoxy-1,4-para-phenylene vinylene) oligomers. *Chem. Phys. Lett.* **223**, 82–88 (1994).
- Gartstein, Y. N., Rice, M. J. & Conwell, E. M. Electron–hole interaction effects in the absorption spectra of phenylene-based conjugated polymers. *Phys. Rev. B* **52**, 1683–1691 (1995).
- Chandross, M. *et al.* Optical absorption in the substituted phenylene-based conjugated polymers: Theory and experiment. *Phys. Rev. B* **55**, 1486–1496 (1997).
- Kersting, R. *et al.* Femtosecond energy relaxation in π -conjugated polymers. *Phys. Rev. Lett.* **70**, 3820–3823 (1993).
- Hayes, G. H., Samuel, I. D. W. & Phillips, R. T. Exciton dynamics in electroluminescent polymers studied by femtosecond time-resolved photoluminescence spectroscopy. *Phys. Rev. B* **52**, R11569–R11572 (1995).
- Harrison, N. T., Hayes, G. R., Phillips, R. T. & Friend, R. H. Singlet intrachain exciton generation and decay in poly(p-phenylenevinylene). *Phys. Rev. Lett.* **77**, 1881–1884 (1996).
- Rothberg, L. J. *et al.* Intrinsic and extrinsic constraints on phenylenevinylene polymer electroluminescence. *Synth. Met.* **78**, 231–236 (1996).
- Barth, S., Bässler, H., Rost, H. & Hörhold, H. H. Extrinsic and intrinsic dc photoconductivity in a conjugated polymer. *Phys. Rev. B* **56**, 3844–3851 (1997).
- Pope, M. & Swenberg, C. E. *Electron Processes in Organic Crystals* (Clarendon, Oxford, 1982).
- Gutmann, F. & Lyons, L. E. *Organic Semiconductors* (Wiley, New York, 1967).
- Salaneck, W. R., Stafström, S. & Brédas, J.-L. *Conjugated Polymer Surfaces and Interfaces* (Cambridge Univ. Press, 1996).

Acknowledgements. We thank M. Harrison, S. Graham, P. Hamer and S. C. Moratti for discussions, and J. J. M. Halls for assistance with the UV photovoltaics equipment and for discussions. One of us (A.K.) thanks Wolfson College and Peterhouse in Cambridge for financial support. The Cambridge–Mons collaboration is supported by the European Commission (TMR Network ‘SELOA’ and ESPRIT project ‘LEDPOS’). Work in Mons is partly supported by the Belgian Prime Minister Services for Scientific, Technical, and Cultural Affairs (IAP in Supramolecular Chemistry and Catalysis), FNRS-FRFC, and an IBM Academic Joint Study. D.B. is Chargé de Recherches FNRS.

Correspondence and requests for materials should be addressed to A.K. (e-mail: ak10007@cus.cam.ac.uk).

Current-limiting mechanisms in individual filaments extracted from superconducting tapes

X. Y. Cai*, A. Polyanskii†, Q. Li‡, G. N. Riley Jr‡ & D. C. Larbalestier*

* Applied Superconductivity Center, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

† Institute of Solid State Physics, Russian Academy of Science, 142432 Chernogolovka, Russia

‡ American Superconductor Corporation, Westborough, Massachusetts 01581, USA

Many large-scale applications of high-temperature superconductors depend crucially on the ability to achieve high critical-current densities J_c (of the order of 10^5 A cm^{-2}). Existing silver-sheathed (Bi,Pb)₂Sr₂Ca₂Cu₃O_x (BSCCO) tapes have J_c values that come within about 25% of this target^{1–4}, these values being limited by the fact that the supercurrent flows percolatively around barriers that occur over many length scales⁵. To elucidate the nature of these barriers, we have measured the transport properties of individual filaments extracted from very-high- J_c multifilament tapes⁶. We find that J_c for individual filaments reaches at least $8 \times 10^4 \text{ A cm}^{-2}$ —about 50% higher than the average value over the whole cross-section of the wire. Although we injected the current to flow along the crystallographic a – b planes of the material, we found that all filaments possessed local characteristics of c -axis transport, indicating the presence of occasional nanometre- to micrometre-scale barriers at basal-plane-faced grain boundaries. An independent and much larger limiting influence on the critical current comes from unhealed cracks produced by deformation during the processing of the wires. These results provide direct evidence that better processing methods aimed at improving the c -axis alignment and at inhibiting residual cracks should raise the accessible J_c values towards those needed for applications.

The filaments studied were extracted from an 85-filament, rectangular conductor, the J_c (at 77 K, 0 T) value of which was 54 kA cm^{-2} , averaged over the whole BSCCO cross-section. This value is characteristic of the very best conductor samples which can be made reproducibly⁶. The filaments, extracted by dissolving the silver away in a NH₄OH/H₂O solution, were 5–10 μm thick, 100–200 μm wide, and several millimetres long (Fig. 1a). Current leads were attached to both top and bottom faces with silver-epoxy to provide symmetric current feed into the a – b planes, as shown in Fig. 1a inset. Owing to filament burn-up in some of the first tests, sections between the voltage probes ($\sim 0.2 \text{ mm}$ apart) were patterned into $\sim 50\text{-}\mu\text{m}$ -wide bridges by slicing off the edges of the filaments with a laser-cutter (we admit the possibility that this may have removed the very highest- J_c regions of the filaments⁷).

Voltage–current characteristics were measured with a precision of $\sim 10 \text{ nV}$. J_c was defined by I_c/A , where I_c is the critical current defined at the usual $1 \mu\text{V cm}^{-1}$ criterion and A is the minimum cross-sectional area of the filament between the voltage taps. Local variations in A which depend on the plate-like nature of the BSCCO grains produce local variations of the filament thickness of a few micrometres. Both thickness and width were measured with a light microscope to an accuracy of $\sim 1 \mu\text{m}$, making the width uncertain to $\sim 5\%$ and the thickness uncertain to 10–20%. J_c (77 K, 0 T) values obtained from seven different filaments ranged from 30 to 80 kA cm^{-2} , thus showing considerable macroscopic inhomogeneity

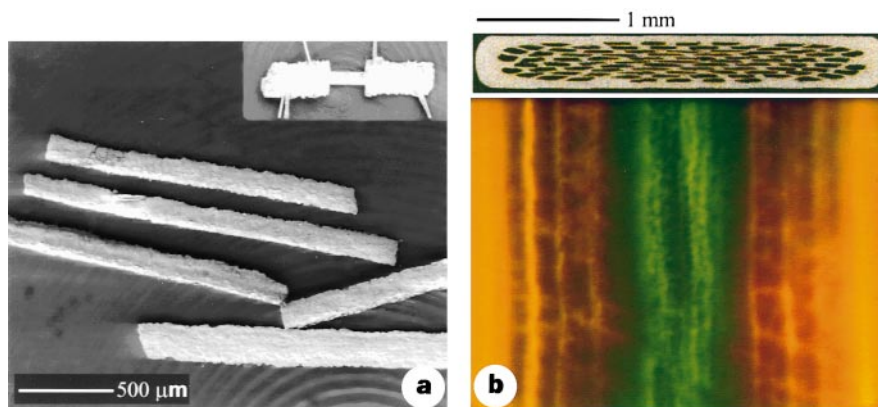


Figure 1 Images of isolated filaments and the whole tape. **a**, Scanning electron micrograph of filaments isolated from a 85-filament Ag-sheathed Bi-2223 tape with self-field J_c (at 77 K) of 54 kA cm^{-2} . Inset, filament with leads attached and with a laser-cut central section to avoid sample burn-up. **b**, Magneto-optic (MO) image of the field distribution for the whole tape. The image was taken by cooling the sample in a field of 40 mT applied perpendicular to the broad face of the tape (that is, approximately parallel to the c -axis) and then switching off the field at 16 K.

of the conductor. Some of the inhomogeneities are obvious in the magneto-optic image of Fig. 1b. A key feature of this image is that flux penetrates intermittently in a direction orthogonal to the filament axis in a manner characteristic of the cracks produced by the intermediate rolling step required to densify the tape after its first heat treatment⁸. Such cracks interfere with longitudinal current flow over large areas. This magneto-optic image supports the idea that the J_c of individual filaments is strongly dependent on local position over scales of $<1 \text{ mm}$ (refs 9, 10), and may be significantly lower for the outer filaments.

Figure 2a shows the temperature dependence of the resistivities ρ_n and ρ_p , where ρ_n and ρ_p were measured with current flowing normal and parallel to the broad filament face, respectively, thus in directions nominally normal to and parallel to the a - b planes of the BSCCO grains. The value of ρ_n/ρ_p at 300 K, 25, is substantially larger than the values of 4–10 reported earlier for lower- J_c (10–20 kA cm^{-2}) single-core tapes, indicating that the grains are much better aligned than is normally the case for monofilament tapes^{11–13}. X-ray rocking curves on a tape from this series showed a c -axis distribution of 13° (full-width at half-maximum), $\sim 50\%$ less than earlier tapes with J_c of $\sim 20 \text{ kA cm}^{-2}$. The resistive transitions measured over six orders of magnitude in different fields are shown in Fig. 2b. A 'foot' structure, characteristic of a weak link, can be clearly seen in the zero-field transition curve, but this is absent in applied fields.

In spite of their excellent properties, a striking result is that all filaments showed superconducting transport behaviour (Fig. 3) which shares multiple, common features with that reported for c -axis transport in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_x$ (Bi-2212) single crystals, even though current was always injected to flow along the filament—that is, in principle, along the a - b planes. The first feature is the precipitous voltage jump in the V - I curve shown in Fig. 3a inset and in Fig. 4. This voltage jump is similar to that reported for c -axis transport in Bi-2212 single crystals^{14–16}, where the jump was regarded as the sum of the individual gap voltages $I_c(c) \times R_n$ generated by current flow across each Sr-O/Bi-O/Bi-O/Sr-O blocking layer, each of which acts as an intrinsic Josephson junction: here $I_c(c)$ is the c -axis junction critical current and R_n is the c -axis resistance. To avoid the possibility that this voltage jump was due to thermal runaway, we measured it from 70 K to T_c ($= 107 \text{ K}$), finding that the jump showed a smooth monotonic decline, consistent with the temperature dependence of the energy gap.

Darker regions show strong superconductivity, while lighter regions are those of easy flux entry and escape. Comparison of the MO image to the filament cross-section above shows vertical bands that indicate the filament structure of the tape, while residual unhealed cracks in outer filaments are revealed by the quasi-periodic horizontal lighter lines crossing the filaments. We note that the MO image is a plan view of the tape, while the upper microstructural image is a cross-section.

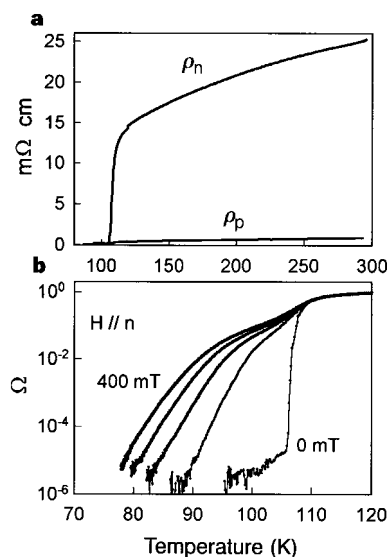


Figure 2 Resistance versus temperature measurements. **a**, Resistivity versus temperature curves of a high- J_c filament measured with the current flowing parallel to (ρ_p) and normal to (ρ_n) the broad face of the filament. **b**, Detailed resistance versus temperature curves measured in fields of 0, 100, 200, 300 and 400 mT.

The second feature is that V - I curves measured at high fields showed ohmic behaviour over multiple orders of magnitude of voltage (Fig. 3a). This same feature was reported as being characteristic of quasi-particle conduction across each blocking layer^{15,17}.

The third feature is the anomalous increase of I_c from $\sim 10 \text{ K}$ to 25 K measured in strong fields applied perpendicular to the broad filament face, that is, approximately parallel to the c -axis, as shown in Fig. 3b. The same feature was reported by Rodriguez *et al.*¹⁸ and Cho *et al.*¹⁵ for c -axis transport of Bi-2212 single crystals with field parallel to the c -axis. They proposed that the c -axis critical current I_c measured in field decreases below 25 K because the rapid increase of flux pinning centre strength at lower temperatures distorts the vortex positions, disrupting phase coherence across the blocking

layers and thus decreasing I_c (c). Above ~ 25 K, $I_c(c)$ decreases because of the reduction of the energy gap.

These three striking similarities in the nominal a - b plane transport behaviour of even the very highest- J_c filaments to those explicitly measured for c -axis transport in Bi-2212 single crystals convince us that small regions of c -axis current flow interrupt the largely a - b plane current path at least once between voltage taps. The length of the c -axis current paths can be estimated from the magnitude of the voltage jump. Assuming that the $I_c \times R_n$ product at 77 K (ref. 14) of a single blocking layer (that is, one double Sr-O/Bi-O layer) is ~ 0.5 – 1.5 mV, the voltage jump of ~ 0.5 V of Fig. 3a corresponds to transport across ~ 300 – $1,000$ intrinsic Josephson

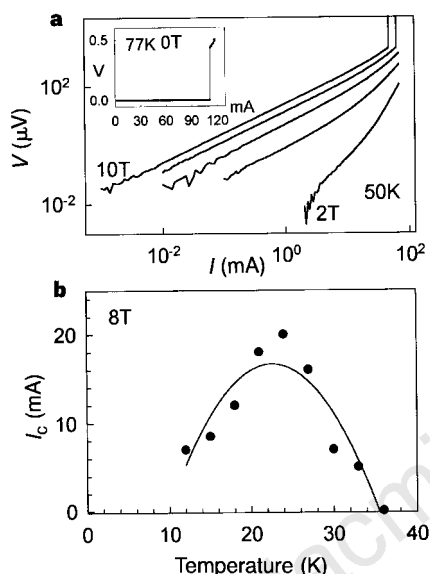


Figure 3 Features of the transport behaviour of high- J_c filaments for currents applied parallel to the filament axis which are similar to those observed for c -axis transport in Bi-2212 single crystals. **a**, V - I curves measured at 2, 4, 6, 8 and 10 T with inset showing the V - I curve measured in self-field. **b**, Temperature dependence of I_c measured at 8 T. All magnetic fields were applied perpendicular to the broad filament face, that is, approximately parallel to the c -axis. These c -axis transport characteristics do not contradict the resistive transition curves shown in Fig. 2, which indicate largely a - b -plane transport for both ρ_p and ρ_n , because the length of the c -axis current path deduced from the characteristics is much smaller (~ 1 μm) than that of the a - b -plane current path (at least 200 μm), as noted in the text.

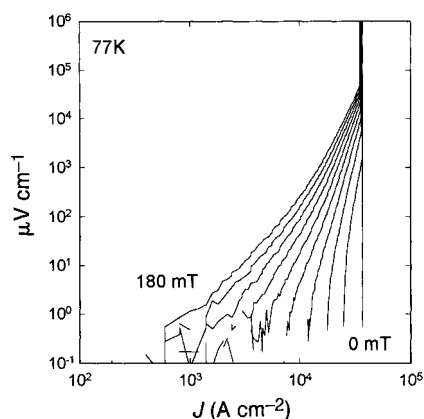


Figure 4 Electric field versus current density curves of a filament measured at 77 K in fields from 0 to 180 mT (in 20 mT steps) applied perpendicular to the broad face.

junctions or a total c -axis path length of ~ 0.5 – 1.5 μm , much less than the spacing between voltage taps (~ 200 μm) or the filament thickness (5 – 10 μm).

As shown in Fig. 3a inset, the filament can maintain a zero-resistance state in self-field at currents $I \leq I_0$, where I_0 is the current of the voltage jump, indicating that the self-field I_c ($= I_0$) is limited by the Josephson junction current of the c -axis current paths. This weak-link behaviour is consistent with the 'foot' structure shown in the zero-field resistive transition of Fig. 2b. Applying external fields larger than ~ 10 – 20 mT produces a change in the current-limiting mechanism, as demonstrated by the changed electric field current density (E - J) characteristics for $J < J_c$, shown in Fig. 4. Such in-field $\log(E)$ versus $\log(J)$ characteristics are widely observed in Ag/Bi-2223 tapes¹⁹ and are generally attributed to depinning of vortex pancakes within the a - b planes.

We attribute our ability to observe c -axis transport characteristics, which were not previously reported for Ag/Bi-2223 tapes, to the important feature that our experiment was conducted on thin, bare filaments, thus avoiding any masking of the intrinsic BSCCO filament characteristics by current shunting into the lower-resistance Ag sheath. We believe that the c -axis transport characteristics arise from the need to intermittently transfer the current from the strongly conducting a - b planes to the poorly conducting c -axis, when the continuous, good a - b plane current path is blocked^{20,21}. We have observed the same ohmic characteristics of Fig. 3a for a 16° [902] Bi-2212 bicrystal, for which the c -axes were misoriented by 11° and for which most of the grain boundary lay along the basal plane of one crystal²². Such basal-plane-faced boundaries are very common in BSCCO tapes, and crossing them requires c -axis current flow. The fact that the c -axis current path length is only of the order of 1 μm may indicate that only a very few boundaries lie within the current path, as the average thickness of BSCCO grains is of this same order. As such boundaries are very common in BSCCO tapes, a low utilization factor of the cross-section for carrying current is naturally suggested.

Taken as a whole, our results show that the critical current of extremely high- J_c single filaments is limited by multiple, independent mechanisms operating on quite different length scales. Thus no single strategy is valid for raising J_c , suggesting both the great potential that still exists for BSCCO and the need to implement clever processing strategies to exploit this potential. At the smallest scale is vortex pinning within the grains. This provides the fundamental upper limit to J_c (Fig. 4) in fields above ~ 10 mT. Recent work shows that the pinning properties and the irreversibility field of Bi-2223 can be significantly improved by controlled processing^{23,24}, although the origin of the enhanced pinning is not clear. The appearance of c -axis Josephson behaviour shows that better c -axis alignment is also highly desirable, as this will diminish the misorientation of the grains, providing better a - b -plane connectivity. Improving the c -axis alignment requires greater attention to filament architecture and to details of the mechanical deformation and reaction processing, tasks which are also vital to healing the cracks that are produced during the deformation step that occurs between heat treatments.

Single-crystal thin films of BSCCO-2223^{25,26} have been reported to have J_c (77 K, 0 T) values of up to 10^6 A cm^{-2} , some 12 times the best present single-filament value. Thus there appears to be considerable scope for using the above strategies to improve BSCCO filament J_c performance by several times. We particularly note the influence of the intermittent cracks suggested by the magneto-optic image of Fig. 1b, as these appear to extend across whole filaments, thus greatly diminishing the local effective filament cross-section. Cracks are particularly hard to account for in any quantitative way, and thus very hard to engineer out of the process. A great advantage of our single, bare-filament experiment is that the influence of cracks is no longer hidden, making it possible to design new processes to eliminate them. $\square$

Received 24 September 1997; accepted 4 March 1998.

1. Leghissa, M. *et al.* Bi-2223 multifilament tapes and multistrand conductors for HTS power transmission cables. *IEEE Trans. Appl. Supercond.* **7**, 355–358 (1997).
2. Hayashi, K. *et al.* Development of Ag-sheathed Bi2223 superconducting wires and their application to magnets. *IEEE Trans. Appl. Supercond.* **7**, 2201–2206 (1997).
3. Marti, F., Grasso, G., Huang, Y. & Flukiger, R. High critical current densities in long lengths of mono- and multifilamentary Ag-sheathed Bi(2223) tapes. *IEEE Trans. Appl. Supercond.* **7**, 2215–2218 (1997).
4. Buzcek, D. M. *et al.* Manufacturing of HTS composite wire for a superconducting power transmission cable demonstration. *IEEE Trans. Appl. Supercond.* **7**, 2196–2200 (1997).
5. Larbalestier, D. C. The road to conductors of high temperature superconductors: 10 years do make a difference! *IEEE Trans. Appl. Supercond.* **7**, 90–97 (1997).
6. Li, Q. *et al.* Progress in superconducting performance of rolled multifilamentary Bi-2223 HTS composite conductors. *IEEE Trans. Appl. Supercond.* **7**, 2026–2029 (1997).
7. Larbalestier, D. C. *et al.* Position-sensitive measurements of the local critical-current density in Ag sheathed high-temperature superconductor (Bi, Pb)₂Sr₂Ca₂Cu₃O₇ tapes. *Physica C* **221**, 299–303 (1994).
8. Parrell, J. A., Polyanski, A. A., Pashitski, A. E. & Larbalestier, D. C. Direct evidence for residual, preferentially-oriented cracks in rolled and pressed Ag-clad BSCCO-2223 tapes and their effect on the critical current density. *Supercond. Sci. Technol.* **9**, 393–398 (1996).
9. Schuster, Th. *et al.* Current capability of filaments depending on their position in (Bi,Pb)₂Sr₂Ca₂Cu₃O₁₀ multifilament tapes. *Appl. Phys. Lett.* **69**, 1954–1956 (1996).
10. Han, Z., Skov-Hansen, P. & Freltoft, T. The mechanical deformation of superconducting BiSrCaCuO/Ag composites. *Supercond. Sci. Technol.* **10**, 371–381 (1997).
11. Cho, J. H. *et al.* Determination of the superconducting current path in Bi2223/Ag tapes. *Appl. Phys. Lett.* **64**, 3030–3032 (1994).
12. Willis, J. O. *et al.* Dependence of the angular anisotropy of the critical current density on texture in silver-sheathed Bi-2223 tapes. *Adv. Cryogen. Eng.* **40**, 9–16 (1994).
13. Thurston, T. R. *et al.* Synchrotron x-ray scattering measurements of bulk structural properties in superconducting (Bi,Pb)₂(Sr₂Ca₂Cu₃O₁₀-Ag) tapes. *J. Appl. Phys.* **79**, 3122–3132 (1996).
14. Kleiner, R. & Muller, P. Intrinsic Josephson effects in high-*T_c* superconductors. *Phys. Rev. B* **49**, 1327–1341 (1994).
15. Cho, J. H., Maley, M. P., Flesher, S., Lacerda, A. & Bulaevskii, L. N. High-magnetic-field transport properties of Bi₂Sr₂CaCu₂O₈ single-crystals. *Phys. Rev. B* **50**, 6493–6496 (1994).
16. Luo, S., Yang, G. & Gough, C. E. Field and temperature-dependence of the *c*-axis critical current in Bi₂Sr₂CaCu₂O₈ single crystals at the melting and 2-dimensional decomposition fields. *Phys. Rev. B* **51**, 6655–6662 (1995).
17. Briceno, G., Crommie, M. F. & Zettl, A. Giant out-of-plane magnetoresistance in Bi-Sr-Ca-Cu-O—a new dissipation mechanism in copper-oxide superconductors. *Phys. Rev. Lett.* **66**, 2164–2167 (1991).
18. Rodriguez, E., Goffman, M. F., Arribere, A., de la Cruz, F. & Schneemeyer, L. F. Critical-state of the weakly coupled 2-dimensional superconductor Bi₂Sr₂CaCu₂O₈. *Phys. Rev. Lett.* **71**, 3375–3378 (1993).
19. Li, Q., Wiesmann, H. J., Suenaga, M., Motowidlo, L. & Haldar, P. Observation of vortex-glass-to-liquid transition in the high-*T_c* superconductor Bi₂Sr₂CaCu₂O₁₀. *Phys. Rev. B* **50**, 4256–4259 (1994).
20. Feng, Y. & Larbalestier, D. C. Texture relationships and interface structure in Ag-sheathed Bi(Pb)-Sr-Ca-Cu-O superconducting tapes. *Interface Sci.* **1**, 401–410 (1994).
21. Eibl, O. The high-*T_c* compound (Bi,Pb)₂(Sr₂Ca₂Cu₃O₁₀): Features of the structure and microstructure relevant for devices in magnet and energy technology. *Supercond. Sci. Technol.* **8**, 833–861 (1995).
22. Larbalestier, D. C. *et al.* in *Proc. European Conf. on Applied Superconductivity (EUCAS)* 29–34 (Conf. Proc. 148, Inst. Phys., Bristol, 1995).
23. Parrell, J. A. *et al.* Enhancement of the 77 K irreversibility field and critical current density of (Bi,Pb)₂Sr₂Ca₂Cu₃O₈ tapes by manipulation of the final cooling rate. *Appl. Phys. Lett.* **69**, 2915–2917 (1996).
24. Anderson, J. W., Dorris, S. E., Parrell, J. A. & Larbalestier, D. C. The effect of lead content on the critical current density, irreversibility field, and microstructure of Ag-clad Bi_{1.8}Pb_{0.2}Sr₂Ca₂Cu₃O₇ tapes. *J. Mater. Res.* (submitted).
25. Hakuraku, Y. & Mori, Z. Rapid annealing effect in the superconducting 2223 Bi(Pb)SrCaCuO thin-films prepared by sputtering. *J. Appl. Phys.* **73**, 309–315 (1993).
26. Wagner, P., Frey, U., Hillmer, F. & Adrian, H. Evidence for a vortex-liquid vortex-glass transition in epitaxial Bi₂Sr₂Ca₂Cu₃O₁₀ thin-films. *Phys. Rev. B* **51**, 1206–1212 (1995).

Acknowledgements. We thank J. Anderson, E. Hellstrom, A. Gurevich, J. Parrell, A. Pashitski and Y. Wu for experimental help and discussions. This work was supported by EPRI and NSF-funded MRSEC.

Correspondence and requests for materials should be addressed to D.C.L. (e-mail: larbales@engr.wisc.edu).

Determination of the absolute chirality of individual adsorbed molecules using the scanning tunnelling microscope

G. P. Lopinski, D. J. Moffatt, D. D. M. Wayner & R. A. Wolkow

Stacie Institute for Molecular Sciences, National Research Council,
100 Sussex Drive, Ottawa, Canada K1A 0R6

The adsorption of organic molecules on solid substrates is a fundamental step in many important heterogeneous catalytic processes, and is also becoming an increasingly significant aspect of surface modification in microelectronics and sensing technology. The conformation of adsorbed molecules not only influences the outcome of surface-catalysed reactions but also

becomes important for recognition processes involved in chemical sensors. The scanning tunnelling microscope (STM) is uniquely able to monitor surface structures at the individual-molecule level¹, and has been shown previously to be capable of distinguishing between different molecular conformations on a surface². Here we show that the geometric configuration (*cis* or *trans*) of several simple alkenes chemisorbed on the silicon (100) surface can be determined using the STM, through its ability to identify individual methyl groups. Because both the position and the orientation of these groups can be seen, we can determine the absolute configuration (*R* or *S*) for each of the chiral centres formed on chemisorption. Thus the STM can probe enantioselective processes at surfaces at the single-molecule level, and may assist in the development of structured chiral surfaces capable of complex recognition tasks.

We have studied the (100) crystal face of silicon, the surface generally used in silicon-based microelectronic devices. The Si(100) surface has been extensively investigated by STM and other techniques³. The salient feature of this surface is the silicon dimer. Each surface atom has two 'back bonds' to the substrate below, shares in one dimer bond, and has one unsatisfied or 'dangling' bond. The dimers form rows that appear as bars in STM images. Alkenes adsorb molecularly on the silicon dimers forming two Si–C bonds with the dangling bonds on each dimer^{4–10}. Our measurements were performed in an ultra-high-vacuum chamber (with a base pressure of 5×10^{-11} torr) which allows the crystal to be kept largely free of contaminants. The n-type (As doped, $0.005 \Omega \text{ cm}$) Si(100) wafers were resistively heated to 680 °C for 12 h to degas the sample and holder, followed by heating to 1,250 °C for several seconds to obtain a clean surface. Adsorption of the subject molecules was achieved by controllably leaking the gaseous molecules into the vacuum chamber.

Figure 1a, b and c show respectively STM images of propylene, *trans*-2-butene and *cis*-2-butene on Si(100). In each case, small protrusions are observed that we associate with the adsorbed molecules. Whereas in the case of propylene the protrusions are randomly distributed, the butenes show what are clearly paired protrusions. These bright features can be associated with the methyl groups within each of these molecules. It is also apparent from Fig. 1 that the images of the *cis* and *trans* molecules are different. The methyl groups of the *cis* isomer define a line that is at a right angle to the dimer rows. In contrast, the methyl units of the *trans* isomer define a line that is angled at $\sim 30^\circ$ to the dimer row. We note that in the particular imaging mode used here, the dimer rows give rise to bars centred between the actual dimer rows. These images probe unoccupied electronic levels, revealing antibonding states which have a node at the dimer centre. The images are therefore consistent with the adsorbed molecule being bound to a single Si dimer.

Assignment of the bright protrusions to methyl groups is further justified by consideration of the bonding geometries of these alkenes on the Si(100) surface, illustrated in the insets of Fig. 1. Calculated adsorption geometries were obtained using the AM-1^{11–15} semi-empirical method and a two-dimer, 31-Si-atom cluster to model the surface (R.A.W. and G.P.L., manuscript in preparation). The alkenes become alkane-like on adsorption, as each carbon atom rehybridizes from sp^2 to sp^3 forming a bond to one silicon atom (of the same dimer) thereby satisfying the dimer dangling bonds. The Si–Si dimer bond lengths only slightly (from 2.37 to 2.38 Å) on alkene adsorption. The C–C bond, originally 1.33 Å in the free molecule, is calculated to increase to 1.50 Å on adsorption, consistent with a change from a double to a single bond in agreement with previous theoretical studies of ethylene/Si(100) (refs 8, 9). The positions of the methyl groups in the images are in quantitative agreement with the calculated bonding geometries. By registering the methyl protrusions of adsorbed propylene with the underlying Si dimers, the methyl group is seen to be displaced ~ 1.1 Å from the centre of these units, in good agreement with the inset of Fig. 1a. Likewise the

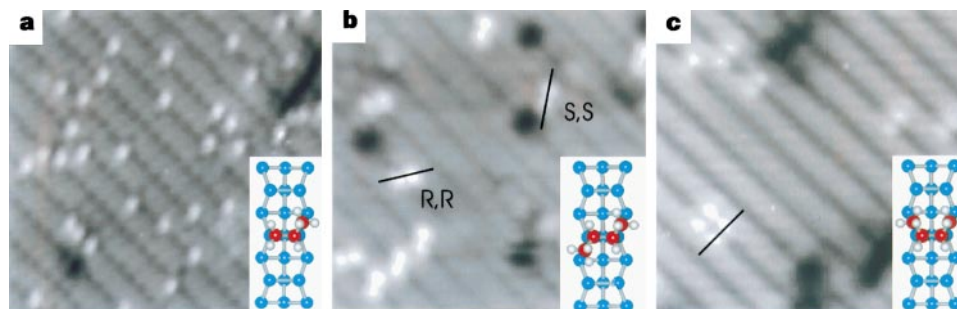


Figure 1 STM images of propylene (**a**), *trans*-2-butene (**b**) and *cis*-2-butene (**c**) on Si(100). The bar-like structures are the dimer rows of the silicon substrate. (The dark regions are due to pre-existing defects.) The white protrusions are due to the methyl groups of the adsorbed molecules. In **b** and **c**, lines have been drawn through the methyl groups of selected molecules to show the orientation of these

with respect to the underlying lattice. In **b**, an example of each enantiomer (*RR* or *SS*) of adsorbed *trans*-2-butene is labelled. The insets show the configuration of each of the alkenes studied on the Si(100) surface. (The dimer rows of the images are rotated $\sim 45^\circ$ relative to the insets.) Image area $75 \times 75 \text{ \AA}$, sample bias +2 V, current 100 pA.

distance between the paired protrusions of the *trans*-2-butene is measured to be $3.4 \pm 0.2 \text{ \AA}$, as compared with $3.6 \text{ \AA}$ obtained from calculation.

The observation of methyl groups as maxima in the STM images is somewhat surprising. Most adsorbates on silicon surfaces appear dark (topographic depressions) because surface-adsorbate bond formation eliminates the dangling bonds that give rise to image features associated with surface silicon atoms¹⁶. In order for a feature to appear bright in an STM image it is generally not sufficient for it to be a topographic protrusion on the surface; it must also have electronic states that are energetically accessible for tunnelling to (or from) the STM tip. Images of hydrogen atom-terminated dimers, for example, reveal a $1.5\text{-\AA}$ depression on the surface relative to unreacted dimers¹⁷. Adsorbed ethylene molecules also appear depressed ($0.1\text{--}0.2 \text{ \AA}$) relative to unreacted dimers¹⁸. The methyl groups of the adsorbed alkenes, being fully saturated, might also be expected to appear as depressions; but *ab initio* calculations using a Si cluster to model the surface indicate that the adsorbed molecules are indeed expected to be visible in STM, in accord with our observations (R.A.W. and G.P.L., manuscript in preparation).

In addition to testing the capacity of STM to determine the geometric configuration of the adsorbed molecule, the alkenes of this study provide an opportunity to monitor the degree of stereoselectivity of the reaction. Previously, *cis*- and *trans*-2-butene have been shown to have different desorption energies from Si(100), demonstrating that the isomers retain their configuration on adsorption¹⁰. Stereoselectivity in this class of reactions has also been suggested by an infrared study of dideuteroethylene/Si(100), relying on matching very small calculated frequency shifts with experiment to determine the dominant adsorbed isomer¹⁹. These methods provide an average measurement from a macroscopic sample. However, the present drive to develop nanoscale devices may require the capability to determine the stereochemistry of extraordinarily small amounts of material with a high degree of precision. STM, with its molecule-by-molecule inspection capability, allows us to determine the degree to which a reaction is stereoselective to arbitrary accuracy (simply determined by the number of molecules inspected). In the present case, inspection of ~ 500 adsorbed molecules revealed $2.1 \pm 0.7\%$ of the *cis* isomer and $1.2 \pm 0.5\%$ *trans* isomer in the *trans* and *cis* samples, respectively. As these fractions are consistent with the specified isomeric impurity level in the gases used, it is evident that no measurable isomerization occurs on adsorption.

The observation that butene adsorption is highly stereoselective places important constraints on the reaction mechanisms. The retention of the stereochemistry of the reactants during addition to the surface implies either a concerted reaction or a stepwise

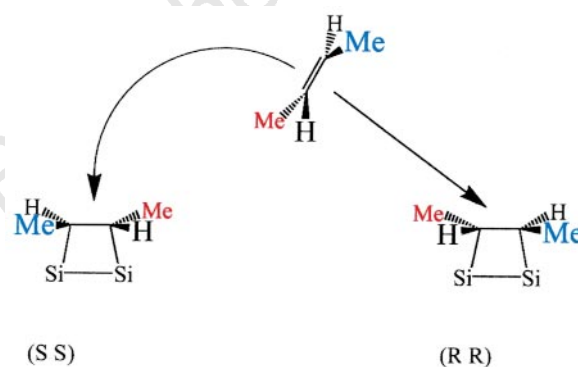


Figure 2 The planar *trans*-2-butene molecule has two distinct faces. Reaction at the *Si* face leads to the *SS* structure shown at the left, whereas reaction at the *Re* face leads to the *RR* structure at the right. Although indistinguishable using optical and other spectroscopic techniques, the individual enantiomers are clearly identified by STM.

process in which the time between the formation of the two Si–C bonds is insufficient to allow a C–C bond rotation. A concerted suprafacial reaction between ethylene and Si(100) has been suggested to be symmetry-forbidden owing to the Woodward–Hoffmann rules¹⁹. However, we note that the Si dimers are in fact tilted (one end up, the other down²⁰) at room temperature, which is probably sufficient to lift this restriction. This tilting, however, is also likely to favour a stepwise mechanism.

There is a further and unique benefit to direct observation of the geometric isomers of single adsorbed molecules, as illustrated by the present case. Knowledge of the spatial relationship between methyl groups of a single reacted alkene provides an opportunity to determine the absolute configuration (*R* or *S*) for each newly formed asymmetric (chiral) centre. As shown in Fig. 2, an unsaturated bond such as that in *trans*-2-butene has two reactive faces, associated with two distinct products. We call these the *Re* and the *Si* faces; the *Re* face leading to a product with *R* chirality, and the *Si* face leading to a product with *S* chirality. Thus the reaction of *trans*-2-butene leads to two chiral products (either *RR* or *SS*) whereas the reaction of *cis*-2-butene results in a non-chiral product (*RS* or *SR*). As the *Re* and *Si* faces of *trans*-2-butene react with equal probability, an equal number of *RR* and *SS* enantiomers are formed. Nevertheless, it is clear from Fig. 1 that *RR* is distinguishable from *SS*. Similarly, by registering the methyl groups with the underlying lattice, the chiralities of adsorbed propylene (*R* or *S*) and *cis*-2-butene (*RS* or *SR*) can be determined. In this way, we can determine

the absolute chirality of individual adsorbed molecules. Although the *Re* and *Si* faces of *trans*-2-butene are identical, the introduction of a single chiral centre adjacent to the double bond removes this degeneracy. We expect that steric inhibition at one of the faces would result in a highly enantioselective process (asymmetric induction). The work reported here shows that it should be possible to determine quantitatively the enantioselectivities of these reactions. □

Received 15 October 1997; accepted 16 February 1998.

1. Binnig, G., Rohrer, H., Gerber, Ch. & Weibel, E. Surface studies by scanning tunneling microscopy. *Phys. Rev. Lett.* **49**, 57–60 (1982).
2. Jung, T. A., Schlitter, R. R. & Gimzewski, J. K. Conformational identification of individual adsorbed molecules with the STM. *Nature* **386**, 696–398 (1997).
3. Becker, R. & Wolkow, R. A. in *Scanning Tunneling Microscopy* (eds Stroscio, J. A. & Kaiser, W. J.) Vol. 27 in *Methods of Experimental Physics* (Academic, 1993).
4. Bozack, M. J., Taylor, P. A., Choyke, W. J. & Yates, J. T. Jr Chemical activity of the C=C double bond. *Surf. Sci.* **177**, 1933–1937 (1986).
5. Yoshinobu, J., Tsuda, H., Onchi, M. & Nishijima, M. The adsorbed states of ethylene on Si(100)(4×2), Si(100)(2×1), and vicinal Si(100)9°: electron energy loss spectroscopy and low energy electron diffraction studies. *J. Chem. Phys.* **87**, 7332–7340 (1987).
6. Clemen, L. et al. Adsorption and thermal behavior of ethylene on Si(100)-(2×1). *Surf. Sci.* **268**, 205–216 (1992).
7. Huang, C., Widdra, W. & Weinberg, W. H. Adsorption of ethylene on the Si(100)-2×1 surface. *Surf. Sci.* **315**, 1953–1958 (1994).
8. Fisher, A. J., Blochl, P. E. & Briggs, G. A. D. Hydrocarbon adsorption on Si(100): when does the silicon dimer bond break? *Surf. Sci.* **374**, 298–305 (1997).
9. Pan, W., Zhu, T. & Yang, W. First-principles study of the structure and electronic properties of ethylene adsorption on Si(100)-2×1 surface. *J. Chem. Phys.* **107**, 3981–3985 (1997).
10. Kiskinova, M. & Yates, J. T. Observation of steric conformation effects in hydrocarbon adsorption and decomposition: cis- and trans-butene-2 on Si(100)-(2×1). *Surf. Sci.* **325**, 1–10 (1995).
11. Dewar, M. J. S. & Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **99**, 4899–4907 (1977).
12. Davis, L. P. et al. MNDO calculations for compounds containing aluminum and boron. *J. Comput. Chem.* **2**, 433–438 (1981).
13. Dewar, M. J. S., McKee, M. L. & Rzepa, H. S. MNDO parameters for third period elements. *J. Am. Chem. Soc.* **100**, 3607 (1978).
14. Dewar, M. J. S., Zoebisch, E. G. & Healy, E. F. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
15. Dewar, M. J. S. & Reynolds, C. H. An improved set of MNDO parameters for sulfur. *J. Comput. Chem.* **2**, 140–148 (1986).
16. Wolkow, R. A. & Avouris, Ph. Atom-resolved surface chemistry using scanning tunneling microscopy. *Phys. Rev. Lett.* **60**, 1049–1052 (1988).
17. Boland, J. J. Role of bond strain in the chemistry of H on Si(100). *Surf. Sci.* **261**, 17–25 (1992).
18. Mayne, A. J. et al. An STM study of the chemisorption of C₂H₄ on Si(001)(2×1). *Surf. Sci.* **284**, 247–256 (1993).
19. Liu, H. & Hamers, R. J. Stereoselectivity in molecule surface reactions: adsorption of ethylene on the Si(100) surface. *J. Am. Chem. Soc.* **119**, 7593–7594 (1997).
20. Wolkow, R. A. Direct observation of an increase in buckled dimers on Si(001) at low temperature. *Phys. Rev. Lett.* **68**, 2636–2639 (1992).

Correspondence and requests for materials should be addressed to R.A.W. (e-mail: bob.wolkow@nrc.ca).

Extremely acid Permian lakes and ground waters in North America

Kathleen Counter Benison^{*†‡}, Robert H. Goldstein^{*}, Brigitte Wopenka[§], Robert C. Burruss^{||} & Jill Dill Pasteris[§]

^{*} Department of Geology, The University of Kansas, Lawrence, Kansas 66045, USA

[†] Kansas Geological Survey, Lawrence, Kansas 66045, USA

[§] Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA

^{||} US Geological Survey, Reston, Virginia 20191, USA

Evaporites hosted by red beds (red shales and sandstones), some 275–265 million years old, extend over a large area of the North American mid-continent¹. They were deposited in non-marine saline lakes, pans and mud-flats², settings that are typically assumed to have been alkaline. Here we use laser Raman microprobe analyses of fluid inclusions trapped in halites from these

Permian deposits to argue for the existence of highly acidic (pH <1) lakes and ground waters. These extremely acidic systems may have extended over an area of 200,000 km². Modern analogues of such systems may be natural acid lake and ground-water systems (pH ~2–4) in southern Australia^{3–9}. Both the ancient and modern acid systems are characterized by closed drainage, arid climate, low acid-neutralizing capacity, and the oxidation of minerals such as pyrite to generate acidity. The discovery of widespread ancient acid lake and groundwater systems demands a re-evaluation of reconstructions of surface conditions of the past, and further investigations of the geochemistry and ecology of acid systems in general.

For this study, halites from three subsurface cores of the Opeche Shale of North Dakota and the Nippewalla Group of Kansas were sampled (Fig. 1). Detailed petrography of the three cores shows that these rocks were deposited in continental saline pan systems dominated by ephemeral saline lakes, saline mud-flats and dry mud-flats². Bedded halites are composed of chevron and cumulate crystals. Chevrons are centimetre-scale, vertically orientated, elongate halite crystals that are interpreted to have grown upwards from the sediment–water interface. Cumulates are millimetre-scale, randomly orientated, equant halite crystals that probably grew at the water surface before sinking down through the water column. The presence of both chevrons and cumulates in halite beds suggests deposition in shallow, non-stratified saline water. Cracked microcrystalline salt crusts indicate desiccation. Dissolution surfaces and halite-filled dissolution pipes indicate dissolution from flooding events. Halite beds are associated with red-bed siliciclastic mudstone and siltstone containing halite-filled vertical and horizontal desiccation cracks and root features. Randomly dispersed in some of the red beds are blocky, clear halite crystals interpreted to have grown displacively in soft sediment from saline ground water. These observations indicate that the red beds were deposited in saline and dry mud-flats surrounding ephemeral saline lakes. The sedimentology of these mid-Permian red beds and evaporites fits the criteria for a saline pan depositional system¹⁰.

Fluid inclusions in both the Opeche and Nippewalla bedded halites and displacive halites are considered to be trapped remnants of Permian lake waters and ground waters, respectively. Primary fluid inclusions are aligned along growth bands in chevron and

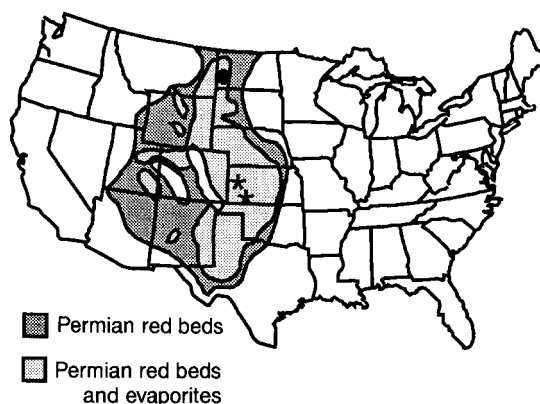


Figure 1 Map of the United States showing the distribution of Permian red beds and evaporites. Approximate locations of cores analysed in this work from the Opeche Shale (filled circle) and the Nippewalla Group (stars) are shown. Opeche Shale samples come from the Gulf-Romanysyn 2-33-4B core (depths of 2,343 m and 2,247 m) of Billings County in southwestern North Dakota. Nippewalla Group samples are from the Anadarko-Davis No. 1 core (depths of 433–403 m) of Seward County in southwestern Kansas and from the Atomic Energy Commission No. 5 core (depths of 632–473 m) of Wichita County in west-central Kansas. Modified from Walker¹.

[‡] Present address: Department of Geology, Central Michigan University, Mt Pleasant, Michigan 48859, USA.

cumulate crystals. They range in size from 1 μm to $\sim 200 \mu\text{m}$, but most are between 5 and 30 μm . Most inclusions are filled with liquid, but some contain a vapour phase and crystals of minerals that were trapped accidentally during halite growth. The origins of some other inclusions, which are isolated in displacive halite crystals, are ambiguous but are likely to represent saline Permian ground waters, rather than lake waters.

Fluid and solid inclusions were analysed non-destructively *in situ* in unmounted pieces of halite using the model S-3000 Jobin-Yvon laser Raman microprobe at Washington University. Raman spectra provide information about the covalently bonded species in fluids and the structure of covalently bonded solids. The peak positions indicate both the speciation of the fluids and the identity of most crystals, whereas peak intensities indicate the relative concentrations of dissolved species. (Knowledge of the concentrations of all species in a solution such as a brine inclusion would permit calculation of its pH. However, only covalently bonded species are Raman-active.) In the inclusions within the Opeche and Nippewalla halite, we detected no dissolved gases, carbonate or bicarbonate species. However, we did find abundant dissolved sulphate, and, in addition, detected the acidic sulphur complex HSO_4^- (see Fig. 2).

The fact that SO_4^{2-} and HSO_4^- can easily be distinguished by Raman spectroscopy^{11–13} means that their relative concentrations can be used to estimate pH (ref. 14). But the pH of the solution is a complex function of both the $\text{HSO}_4^- : \text{SO}_4^{2-}$ ratio and the bulk chemistry of the solution. We therefore spectroscopically analysed an initial set of standard solutions of varying known molar concentrations in glass capillaries, and in synthetic fluid inclusions in halite, within the following relevant chemical systems: (1) $\text{H}_2\text{SO}_4\text{--H}_2\text{O}$, (2) $\text{Na}_2\text{SO}_4\text{--H}_2\text{O}$, (3) $\text{NaCl--H}_2\text{SO}_4\text{--H}_2\text{O}$, (4) $\text{Na}_2\text{SO}_4\text{--H}_2\text{SO}_4\text{--H}_2\text{O}$ and (5) $\text{NaCl--Na}_2\text{SO}_4\text{--H}_2\text{SO}_4\text{--H}_2\text{O}$. The pH of these solutions were determined with short-range pH paper, a pH meter and the PHRQPITZ geochemical model¹⁵. The standard solutions ranged from pH -1.5 to pH 7, and the total dissolved sulphate concentrations in the standards ranged from 2,000 to 20,000 p.p.m. All of the standard solutions yielded Raman spectra with a peak at Raman shift 986 cm^{-1} due to the ν_1 stretching vibration of SO_4^{2-} . The pH of the solution, however, controlled the presence of the two possible peaks for HSO_4^- . For all standard solutions with pH < 0 , there were two HSO_4^- peaks, one at a shift of $\sim 892 \text{ cm}^{-1}$ (due to the S(–OH) stretching vibration) and one at $\sim 1,054 \text{ cm}^{-1}$ (due to the symmetric stretching vibration of SO_3)^{16,17}. In contrast, standard solutions with pH values between 0 and 1 showed only the single HSO_4^- peak at $\sim 1,054 \text{ cm}^{-1}$. Our analyses of standard solutions show that as long as the concentrations of the sulphate species are above the detection limit of the laser Raman microprobe, the presence of the $\sim 892 \text{ cm}^{-1}$ HSO_4^- peak is a function only of pH. All standard solutions with pH > 1 had no detectable HSO_4^- peaks. Thus, the detection of one or both HSO_4^- peaks in the Raman spectrum provides a measure of pH in aqueous acidic sulphate solutions, independent of the presence or absence of neutral salts (NaCl, Na_2SO_4) in the solutions. In addition, all standards and all natural fluid inclusions analysed showed two broad Raman peaks at $\sim 1,645 \text{ cm}^{-1}$ and $\sim 3,500 \text{ cm}^{-1}$ due respectively to the H–O–H bending and O–H stretching vibrational modes of water.

We analysed 92 fluid inclusions in Opeche halite from nine depths within one core. The sampled population includes 83 primary inclusions from seven chevron or cumulate bedded halites and nine isolated inclusions from two beds with displacive halite. Primary fluid inclusions within single chevron growth bands have similar Raman spectra, suggesting constant brine composition at the time of halite precipitation. All 92 fluid inclusions show peaks at $\sim 986 \text{ cm}^{-1}$ for SO_4^{2-} and $\sim 1,054 \text{ cm}^{-1}$ for HSO_4^- . Rare inclusions also have a second weak HSO_4^- peak at $\sim 892 \text{ cm}^{-1}$. The presence of the $\sim 1,054 \text{ cm}^{-1}$ peak for HSO_4^- in every fluid inclusion analysed in the Opeche halite suggests that it was deposited from waters with

pH values below 1, whereas the occurrence of rare inclusions with both HSO_4^- peaks indicates that some halite was deposited from waters with pH < 0 .

The ratio of the $\sim 986 \text{ cm}^{-1}$ SO_4^{2-} and $\sim 1,054 \text{ cm}^{-1}$ HSO_4^- peaks is a simple function of pH only when the concentrations of other ions remain constant. The Raman spectra of the Opeche samples are characterized by a relatively constant ratio of the intensities of the $\sim 986 \text{ cm}^{-1}$ SO_4^{2-} peak and the $\sim 1,054 \text{ cm}^{-1}$ HSO_4^- peak among all samples from each halite bed. Moreover, except for inclusions from the two deepest halite beds, the $\sim 1,054 \text{ cm}^{-1}$ HSO_4^- peak is significantly more intense than the $\sim 986 \text{ cm}^{-1}$ SO_4^{2-} peak (see Fig. 2). In contrast, in standard (NaCl-saturated, $\text{Na}_2\text{SO}_4\text{--H}_2\text{SO}_4\text{--H}_2\text{O}$) solutions that only show one of the HSO_4^- peaks at a pH of 0, and in which sulphuric acid is the sole source of the hydronium ion, the SO_4^{2-} peak is always more intense than the $\sim 1,054 \text{ cm}^{-1}$ HSO_4^- peak. The spectral intensity relation of $\text{HSO}_4^- > \text{SO}_4^{2-}$ in natural fluid inclusions could be reproduced only in an additional set of standard solutions in which the hydronium ion concentration in H_2SO_4 solutions was increased by the addition of HCl (see Fig. 2). This finding suggests that, for the Opeche Shale halite, there was another component of acidity besides H_2SO_4 .

We used the laser Raman microprobe to analyse 51 primary fluid

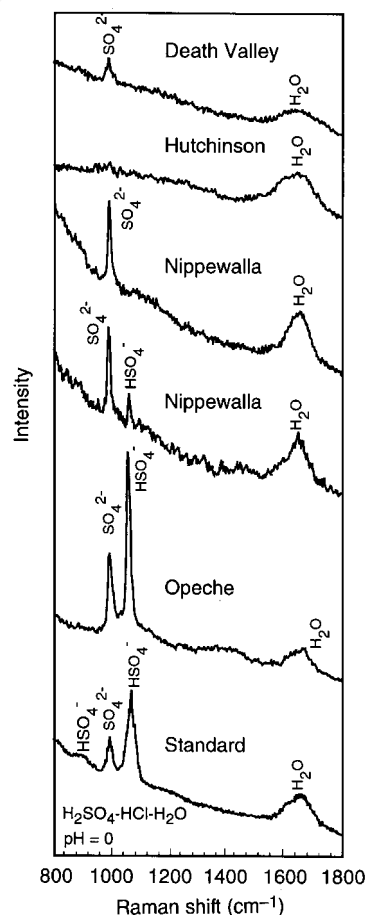


Figure 2 Raman spectra of standard solution and fluid inclusions. Bottom trace, spectrum of an acidic (pH = 0) $\text{NaCl--Na}_2\text{SO}_4\text{--H}_2\text{SO}_4\text{--HCl--H}_2\text{O}$ standard solution in a glass capillary. Other traces are spectra of individual 5–30 μm diameter primary fluid inclusions in Permian halite samples (Opeche Shale, Nippewalla Group and Hutchinson Salt) and modern Death Valley halite. The optical quality of the Nippewalla and Opeche halites yielded high signal to noise ratios, which allow interpretation of the presence and absence of the $\sim 892 \text{ cm}^{-1}$ and $1,054 \text{ cm}^{-1}$ peaks for HSO_4^- . The presence of the $1,054 \text{ cm}^{-1}$ peak indicates that brine inclusions have pH < 1 .

inclusions in Nippewalla chevron and cumulate halite from nine depths in two cores. The liquid phase in all inclusions shows a $\sim 986\text{ cm}^{-1}$ SO_4^{2-} peak. Most of the Nippewalla inclusions have no HSO_4^- peak, indicating that they formed from waters with $\text{pH} > 1$. However, eight of the inclusions have a small $\sim 1,054\text{ cm}^{-1}$ peak for HSO_4^- , suggesting that some of the lake waters at the time of Nippewalla deposition had pH values between 0 and 1.

Bedded halite from other geological settings and ages were analysed to determine how unique HSO_4^- peaks may be within depositional halite. In primary fluid inclusions from the Pennsylvanian Paradox Formation halite, Permian Hutchinson Salt and modern Death Valley halite, no HSO_4^- peaks were detected (see Fig. 2).

Changes in inclusion chemistry after entrapment are highly unlikely, and almost certainly cannot have generated the low pHs measured in the Opeche and Nippewalla fluid inclusions. Evidence for the primary nature of these inclusions and their 'primary acidity' includes (1) the entrapment of inclusions along growth bands, (2) the relative homogeneity of the inclusion compositions, and (3) the chemical difficulty in generating 'secondary acidity' in a fluid inclusion within a sedimentary system. The two types of post-entrapment modification to consider are open-system replacement of early fluids by later, more acidic ones, and closed-system reactions within inclusions that changed their fluid chemistry.

Open-system modification by leakage and refilling of inclusions can be ruled out for three reasons. First, most samples show no petrographical evidence of secondary fractures; second, Nippewalla inclusions retain original depositional temperatures^{18,19}; and third, the Opeche and Nippewalla inclusions are too uniform in composition and pH to have been altered by such a process². Other open-system processes, such as large-scale hydrogen diffusion through halite^{20,21}, are highly unlikely given current knowledge of diffusion into and out of inclusions at temperatures that exist within the sedimentary realm²². Regardless, diffusion of molecular hydrogen should not directly affect pH of the inclusion fluid. Closed-system alteration of inclusion fluids also is highly unlikely. For instance, reactions involving organic acids can be dismissed as the cause of secondary acidity in this study for three reasons. First, laser Raman microprobe analyses would have detected organic compounds if they were abundant within inclusion fluids, but no evidence of organic compounds was found. Second, the low dissociation constants of organic acids cannot cause such extremely low pH. Third, the low iron content in the fluid inclusions² argues against production of H^+ through organic acid degradation facilitated by reduction of Fe^{3+} . In addition, other possible mechanisms of decreasing pH by internal oxidation of species in inclusion fluids, including oxidation of reduced sulphur or reduced iron species, are unlikely because the sedimentological and petrographical evidence for a very shallow-water depositional system, including early oxides and sulphates, indicates a well oxidized setting. We are convinced that the inclusions represent the original Permian lake waters and ground waters, making them the most acidic surface and near-surface waters documented from ancient times.

Extremely acid waters are also known in modern settings, including acid mine-drainage and volcanogenic crater lakes^{23–26}. Acid crater lakes may have $\text{pH} < 0$ caused by outgassing of volcanic H_2S (refs 23, 24). Acid mine-drainage owes much of its acidity (pH as low as -3 ; ref. 25) to weathering of metal sulphide ores. Obviously, these modern settings cannot serve as analogues to the Permian examples but their existence demonstrates that extremely low pHs can be achieved in natural surface waters.

Regional acid systems in southern Australia may serve as a better modern analogue for the Permian examples. Saline lakes and ground waters there are low-pH (~ 2 – 4) sulphuric acid solutions^{3–9}. Like the Permian Opeche Shale and Nippewalla Group, Australian acid waters precipitate evaporite minerals in a carbonate-poor setting with abundant red siliciclastics. Both the Permian and

modern Australian acid systems are situated in tectonically stable, closed-drainage basins characterized by an arid climate and bedrock with little acid-buffering capacity. The initial and most important process leading to acidity in both systems is the oxidation of sulphide minerals^{2–9}. Other acidification processes, which may be mediated by microbes, may help to maintain the acidity of the system. The main difference between the Permian and the modern Australian acid systems is the degree of acidity. The Permian Opeche Shale (with $\text{pH} < 1$), in particular, seems to have pHs much lower than those in the Australian waters ($\text{pH} \sim 2$ – 4). These more extreme pHs may have been caused simply by more sulphide weathering relative to acid-neutralization in the Permian red beds²⁷.

It is clear that the similar tectonic, climatic and geological conditions of the ancient supercontinent Pangaea and modern Australia played an important role in creating the Permian and modern acid lake and groundwater systems. In rocks of other ages from different regions, one might expect that the combination of tectonic stability, internal drainage, arid climate and lack of acid-buffering capacity in the bedrock would produce acid systems, as long as the system was oxidized and there was a source of sulphides.

Thus, ancient acid systems may be important for deciphering palaeoweathering and palaeoclimate trends, as well as having far-reaching implications for understanding chemical evolution of ground waters and red-bed formation. It may be important for the petroleum industry to consider ancient acid systems in its models for porosity formation and occlusion. Furthermore, studies of ancient acid systems may help to predict long-term processes and products in modern acid regimes, information useful in the control and remediation of acid mine-drainage and contaminated waters.

This study has yielded the first, to our knowledge, documentation of ancient extremely acid ($\text{pH} < 1$) lake waters and ground waters, environments that may be more important than previously recognized. The detection of HSO_4^- peaks by laser Raman microprobe spectroscopy in fluid inclusions throughout the Opeche halite and in some of the Nippewalla halite is evidence of ancient extremely acid waters. Based on evidence of significantly long-lived acidity in the northern and southernmost extent of this Permian red-bed-hosted system, and published interpretations of marginal marine evaporites to the south^{28–30}, we estimate that this acid system may have covered an area of $\sim 200,000\text{ km}^2$. This discovery of ancient acid surface and near-surface waters should lead to wider-reaching investigations and re-evaluations of: (1) the ecology and biogeochemistry of ancient lake systems, most of which have been assumed to have been freshwater or saline-alkaline systems; (2) the chemical evolution of lake waters and ground waters, adding the new component of evolution of acidity in the context of weathering and climate; and (3) the origin of both terrestrial and martian red beds, some of which may have been associated with acid systems. □

Received 9 December 1997; accepted 26 January 1998.

- Walker, T. R. Formation of red beds in modern and ancient deserts. *Geol. Soc. Am. Bull.* **78**, 353–368 (1967).
- Benison, K. C. *Acid Water Deposition and Diagenesis of Permian Red Bed-hosted Evaporites, Midcontinent, U.S.A.* Thesis, Univ. Kansas (1997).
- Alpers, C. N., Rye, R. O., Nordstrom, D. K., White, L. D. & King, B. S. Chemical, crystallographic and stable isotope properties of alunite and jarosite from acid-hypersaline Australian lakes. *Chem. Geol.* **96**, 203–206 (1992).
- Hines, M. E., Lyons, W. B., Lent, R. M. & Long, D. T. Sedimentary biogeochemistry of an acidic, saline groundwater discharge zone in Lake Tyrell, Victoria, Australia. *Chem. Geol.* **96**, 53–65 (1992).
- Long, D. T. & Lyons, W. B. Aridity, continental weathering, and groundwater chemistry. *Geol. Soc. Am. Abstr. Progn.* **22**, 295 (1990).
- Long, D. T. & Lyons, W. B. Aridity, continental weathering, and ground-water chemistry. *Geol. Soc. Am. Today* **2**, 185–190 (1992).
- Long, D. T. et al. Geochemistry of acid brines: Lake Tyrell, Victoria, Australia. *Chem. Geol.* **96**, 33–52 (1992).
- Macumber, P. G. Hydrological processes in the Tyrell Basin, southeastern Australia. *Chem. Geol.* **96**, 1–18 (1992).
- McArthur, J. M., Turner, J., Lyons, W. B., Osborn, A. O. & Thirwall, M. F. Hydrochemistry on the Yilgarn Block, Western Australia: Ferrolisis and mineralisation in acidic brines. *Geochim. Cosmochim. Acta* **55**, 1273–1288 (1991).
- Lowenstein, T. K. & Hardie, L. A. Criteria for the recognition of salt-pan evaporites. *Sedimentology* **32**, 627–644 (1985).
- Cunningham, K. M., Goldberg, M. C. & Weiner, E. R. Investigation of detection limits for solutes in water measured by laser Raman spectrometry. *Anal. Chem.* **49**, 70–75 (1977).

12. Moissette, A., Dubessy, J., Cathlineau, M., Ramanaidou, E. & Le Gleuher, M. Fluids with high acidity evidenced by Raman analysis of sulfate and bisulfate ions in fluid inclusions from a gold showing (Brazil). *IAGOD Symp. Progn. Abstr.* A18 (1990).
13. Rosasco, G. J. & Roedder, E. Application of a new Raman microprobe spectrometer to nondestructive analysis of sulfate and other ions in individual phases in fluid inclusions in minerals. *Geochim. Cosmochim. Acta* **43**, 1907–1915 (1979).
14. Dubessy, J., Boiron, M., Moissette, A., Monnin, C. & Stretenskaya, N. Determinations of water, hydrates, and pH of fluid inclusions by micro-Raman spectrometry. *Eur. J. Min.* **4**, 885–894 (1992).
15. Plummer, L. N., Parkhurst, D. L., Fleming, G. W. & Dunkle, S. A. *A Computer Program Incorporating Pitzer's Equations for Calculation of Geochemical Reactions in Brines* (Water Res. Invest. Rep. 88-4153, US Geol. Surv., Reston, VA, 1988).
16. Turner, D. J. Raman spectral study of bisulphate ion hydration. *J. Chem. Soc. Faraday Trans.* **68**, 643–648 (1972).
17. Rudolph, W. Structure and dissociation of the hydrogen sulphate ion in aqueous solution over a broad temperature range: a Raman study. *Z. Phys. Chem.* **194**, 73–95 (1996).
18. Benison, K. C. Surface water paleotemperatures from Permian Nippewalla Group halite, Kansas. *Carbonates Evaporites* **10**, 242–248 (1995).
19. Benison, K. C. & Goldstein, R. H. Permian paleoclimate data from fluid inclusions in halite. *Chem. Geol.* (in the press).
20. Hall, D. L., Bodnar, R. J. & Craig, J. R. Evidence for postentrapment diffusion of hydrogen into peak metamorphic fluid inclusions from the massive sulfide deposits at Ducktown, Tennessee. *Am. Mineral.* **76**, 1344–1355 (1991).
21. Marvoges, J. A. & Bodnar, R. J. Hydrogen movement into and out of fluid inclusions in quartz: Experimental evidence and geologic implications. *Geochim. Cosmochim. Acta* **58**, 141–148 (1994).
22. Goldstein, R. H. & Reynolds, T. J. *Systematics of Fluid Inclusions in Diagenetic Minerals* 199 (Short Course 31, SEPM (Soc. for Sed. Geol.), Tulsa, OK, 1994).
23. Delmelle, P. & Bernard, A. Geochemistry, mineralogy, and chemical modeling of the acid crater lake of Kawak Ijen Volcano, Indonesia. *Geochim. Cosmochim. Acta* **58**, 2445–2460 (1994).
24. Rowe, G. L. *The Acid Crater Lake System of Poas Volcano, Costa Rica: Geochemistry, Hydrology, and Physical Characteristics*. Thesis, Pennsylvania State Univ. (1991).
25. Nordstrom, D. K., Valentine, S. D., Ball, J. D., Plummer, L. N. & Jones, B. F. Measurement of negative pH values and high metal concentrations in extremely acidic mine waters from Iron Mountain, California. *Geol. Soc. Am. Abstr. Progn.* **23**, 383 (1991).
26. Nordstrom, D. K. In *Acid Sulfate Weathering* (eds Kittrick, J. A., Fanning, D. S. & Hossner, L. R.) 37–56 (Spec. Publ. 10, Soil Sci. Soc. Am., Madison, WI, 1982).
27. Berendsen, P. Permian red bed copper mineralization in south-central Kansas. *Acad. Sci. Trans. Abstr.* **83**, 122 (1980).
28. Hovorka, S. Depositional environments of marine-dominated bedded halite, Permian San Andres Formation, Texas. *Sedimentology* **34**, 1029–1054 (1987).
29. Lowenstein, T. K. Origin of depositional cycles in a Permian "saline giant": the Salado (McNutt zone) evaporites of New Mexico. *Geol. Soc. Am. Bull.* **100**, 592–608 (1988).
30. Presley, M. W. Evolution of Permian evaporite basin in Texas panhandle. *Bull. Am. Assoc. Petrol. Geol.* **71**, 167–190 (1987).

Acknowledgements. We thank T. K. Lowenstein and R. J. Bodnar for suggestions that improved this manuscript. This work was partially funded by NSF, GSA and Sigma Xi.

Correspondence and requests for materials should be addressed to K.C.B. (e-mail: kathleen.benison@cmich.edu).

Boreal forest plants take up organic nitrogen

Torgny Näsholm, Alf Ekblad*, Annika Nordin, Reiner Giesler*, Mona Högborg* & Peter Högborg*

Department of Forest Genetics and Plant Physiology and * Section of Soil Science, Department of Forest Ecology, Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden

Plant growth in the boreal forest, the largest terrestrial biome, is generally limited by the availability of nitrogen. The presumed cause of this limitation is slow mineralization of soil organic nitrogen^{1,2}. Here we demonstrate, to our knowledge for the first time, the uptake of organic nitrogen in the field by the trees *Pinus sylvestris* and *Picea abies*, the dwarf shrub *Vaccinium myrtillus* and the grass *Deschampsia flexuosa*. These results show that these plants, irrespective of their different types of root–fungal associations (mycorrhiza), bypass nitrogen mineralization. A trace of the amino acid glycine, labelled with the stable isotopes ¹³C and ¹⁵N, was injected into the organic (mor) layer of an old successional boreal coniferous forest. Ratios of ¹³C: ¹⁵N in the roots showed that at least 91, 64 and 42% of the nitrogen from the absorbed glycine was taken up in intact glycine by the dwarf shrub, the grass and the trees, respectively. Rates of glycine uptake were similar to those of ¹⁵N-ammonium. Our data indicate that organic nitrogen is important for these different plants, even when they are competing with each other and with non-symbiotic microorganisms. This has major implications for our understanding of the

effects of nitrogen deposition, global warming and intensified forestry.

Current conceptual models of nitrogen cycling in boreal forests are based on the assumption that mineralization of organic nitrogen is a prerequisite for plant nitrogen acquisition. Laboratory studies have shown, however, that some plants can use amino acids^{3,4} and proteins⁵, thereby bypassing the common mineralization pathway. This has not been demonstrated in the field, where plants compete with soil microorganisms held to be superior in using organic substrates. It has been proposed that plants forming ericoid mycorrhiza and ectomycorrhiza are especially efficient at using organic nitrogen⁶. This would give them a competitive advantage over plants forming arbuscular mycorrhiza and over non-mycorrhizal plants, and help explain their dominance in boreal forests⁶, in which levels of organic nitrogen are typically high in the soil (compare with Table 1). If so, anthropogenic inorganic nitrogen deposition, climate warming and intensified silviculture could alter the composition of plant communities by promoting a shift from organic- towards inorganic-nitrogen nutrition. Again, field evidence demonstrating that organic-nitrogen uptake occurs and is exclusive to some groups of plants is lacking. The interaction between the abiotic soil system, mycorrhizal fungi, other microorganisms and plants is complex⁷, and studies showing higher gross turnover rates of inorganic nitrogen pools than previously assumed⁸ call for intensified studies of the nitrogen cycle, taking into account higher levels of complexity⁹.

The lack of data showing organic nitrogen uptake by plants in the field relates to methodological problems. Classic ¹⁵N-tracer technique cannot separate uptake of intact amino acids from uptake of mineralized nitrogen. ¹⁴C-labelling has been used in studies of hydroponic (soil-free, nutrient-water) systems⁴, but not in the field. Double-labelled (¹³C, ¹⁵N) glycine was used in a macrocosm study of arctic tundra, showing that plant uptake of glycine-derived nitrogen was as fast as uptake of NH₄⁺, but this gave no evidence that the ¹³C-labelled tracer was taken up¹⁰. Variations in the natural abundance of ¹⁵N among plants in the field may be related to use of different nitrogen sources, but analysis of these variations is unlikely to yield conclusive quantitative evidence of the relative importance of organic and inorganic nitrogen uptake¹¹.

We aimed to test whether uptake of organic nitrogen occurs in the field, and to test whether plants with different types of mycorrhiza differ in this respect. We performed a low-tracer-level study with minimal disturbance of the soil–plant system in a late successional coniferous forest in northern Sweden (Table 1). Double, universally labelled (U) amino acid (U-¹³C₂, ¹⁵N-glycine) injected at 1 kg nitrogen per hectare into the mor layer was used to study the importance of direct uptake of intact amino acid compared with uptake of nitrogen from mineralized amino acid, that is ¹⁵N-NH₄⁺. Typical European boreal forest species¹² (the ectomycorrhizal trees *Pinus sylvestris* L. and *Picea abies* (L.) Karst., the ericoid-mycorrhizal dwarf shrub *Vaccinium myrtillus* L. and the arbuscular-mycorrhizal grass *Deschampsia flexuosa* (L.) Trin.) were included in the study. Sequential sampling of above- and below-ground plant parts and mor layer soil was conducted after 6 h (2 h for above-ground parts), 1 day and 7 days. Respiration of ¹³C-CO₂ was measured after 2 h, 1 day and 7 days.

Table 1 Characteristics of the mor layer at the site studied, Renberget

pH*	3.1 ± 0.0
C:N ratio	36.6 ± 0.9
Total soil nitrogen (kg ha ⁻¹)	302 ± 33
Microbial nitrogen† (kg ha ⁻¹)	27 ± 3
Extractable amino-acid nitrogen‡ (kg ha ⁻¹)	1.0 ± 0.1
Extractable inorganic nitrogen‡ (kg ha ⁻¹)	0.3 ± 0.05

* 0.04 M CaCl₂, soil:solution ratio 1:10.

† Method of ref. 20, using 0.45 as an estimate of the fraction of nitrogen released by fumigation.

‡ In 2 M KCl, soil:solution ratio 1:10.

Data (averages ± 1 s.e.m.) are from control plots (n = 22–24).

The results clearly show that a significant proportion of the supplied tracer was absorbed as intact amino acid. This is illustrated by plots of excess ^{13}C against excess ^{15}N in the soluble nitrogen fraction in roots (Fig. 1), and the highly significant regressions produced ($P < 0.001$). By comparing the slopes of these lines with the slope of 2, derived from the $^{13}\text{C}:^{15}\text{N}$ ratio of the tracer, a conservative estimate of the fraction of nitrogen taken up as amino acid can be obtained. After 6 h, this fraction was at least 42%, 64% and 91% for the trees, the grass and the dwarf shrub, respectively. Corresponding values after 1 day were 45%, 50% and 61%. Seven days after tracer application, there was no clear linkage between ^{13}C and ^{15}N levels in the trees and the grass, although the association was still seen in the dwarf shrub.

The slopes of the regression lines from the first samples show that the dwarf shrub had higher $^{13}\text{C}:^{15}\text{N}$ values than the trees (analysis of covariance followed by Tukey's test; $P < 0.01$) (ref. 13). Tracer ^{15}N was also detected in shoots of the grass after only 2 h and in dwarf shrub shoots after 1 day (Table 2). The ^{15}N -tracer level of both roots (Fig. 1) and shoots (Table 2) was similar for glycine-treated and NH_4^+ -treated plants, giving further support to the hypothesis that glycine-derived nitrogen was, to a large extent, taken up as the intact amino acid¹⁰.

Metabolism of glycine in plants or microorganisms occurs mainly by the route of serine synthesis, whereby 2 mol glycine are converted into 1 mol serine, 1 mol NH_3 and 1 mol CO_2 (ref. 14). Ammonium is re-assimilated, whereas CO_2 is presumably not re-assimilated in non-photosynthetic tissue. We used glycine labelled at both carbon

positions, and the likely loss of the carboxy group during incorporation could explain the very rapid shift in the $^{13}\text{C}:^{15}\text{N}$ ratio from 2:1 to, or towards, 1:1 (Figs 1 and 2). Hence, our estimates of organic nitrogen uptake are probably underestimates, and differences between species could be partly related to differences in rates of decarboxylation. In the previous study of tundra plants¹⁰, only the carboxy carbon of the glycine used was labelled, which may explain the difficulties in detecting the ^{13}C -labelled tracer. Here, some of the variability in natural abundance of ^{13}C relates to differences between plants in shaded and unshaded environments¹⁵. Analysis of the chitin content¹⁶ of the roots indicated that all plants were mycorrhizal, with values of 0.5, 1.1, and 2.2 mg chitin per g root dry weight in the dwarf shrub, the grass and the trees, respectively. These data reflect the large proportion of rhizomes found in the samples of dwarf shrub roots. We cannot deduce how much of the ^{13}C and ^{15}N label was located in either plant or fungal cells. However, as the fungus is an integral part of the mycorrhizal root, this distinction is not critical in the context of biogeochemical nitrogen cycling. The most important consideration is whether or not organic nitrogen is taken up from the soil by the mycorrhizal roots.

Measurements in sifted soil from the labelled plots (Fig. 2) show that $64 \pm 2\%$, that is, most of the nitrogen applied as either glycine or NH_4^+ , was retained in this compartment. Unlike in previous studies, we did not prevent leaching and lateral flow or transport of glycine or NH_4^+ by roots out of the labelled plots, to minimize disturbance to the studied system. The average $^{13}\text{C}:^{15}\text{N}$ ratio for the

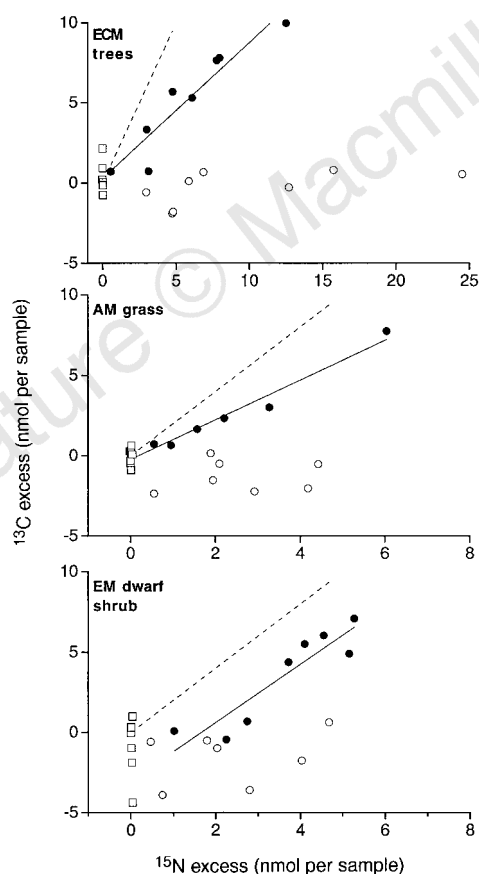


Figure 1 The relationship between excess ^{13}C and excess ^{15}N in the soluble nitrogen fraction in plant roots from the plots 6 h after injecting the mor layer with water (control plots, squares) or solutions containing $(\text{U-}^{13}\text{C}_2, ^{15}\text{N})$ glycine (filled circles) or $^{15}\text{NH}_4^+$ (open circles). The $^{13}\text{C}:^{15}\text{N}$ ratio of the glycine (2:1) injected (broken lines) and the regressions relating to glycine-treated plots (unbroken lines) are shown. ECM, ectomycorrhizal (slope = 0.85, $r^2 = 0.89$); AM, arbuscular mycorrhizal (slope = 1.24, $r^2 = 0.97$); EM, ericoid mycorrhizal (slope = 1.81, $r^2 = 0.84$); $n = 6-8$.

Table 2 $\delta^{15}\text{N}$ of shoots from the dwarf shrub *V. myrtillus* and the grass *D. flexuosa*

Species	Treatment	$\delta^{15}\text{N}$ of shoots*		
		2 h	1 day	7 days
Dwarf shrub	Control	-1.5 ± 0.6	-1.5 ± 0.6	-1.1 ± 0.7
	$^{15}\text{NH}_4^+$	1.5 ± 1.5	2.5 ± 3.4	11.5 ± 7.0
	$(\text{U-}^{13}\text{C}_2, ^{15}\text{N})$	-0.5 ± 0.7	3.3 ± 1.7	3.8 ± 1.4
	Glycine			
Grass	Control	-0.1 ± 0.9	-1.0 ± 0.7	-0.3 ± 0.7
	$^{15}\text{NH}_4^+$	2.1 ± 0.8	12.5 ± 10	64.9 ± 36.5
	$(\text{U-}^{13}\text{C}_2, ^{15}\text{N})$	2.6 ± 1.4	3.3 ± 1.5	20.5 ± 8.2
	Glycine			

* Denotes part per thousand (‰) deviations from the ratio $^{15}\text{N}:^{14}\text{N}$ in atmospheric N_2 . These results were obtained from plots treated with either water (control), $^{15}\text{NH}_4^+$ or $(\text{U-}^{13}\text{C}_2, ^{15}\text{N})$ glycine. (Averages $\pm$ s.e.m., $n = 8$.)

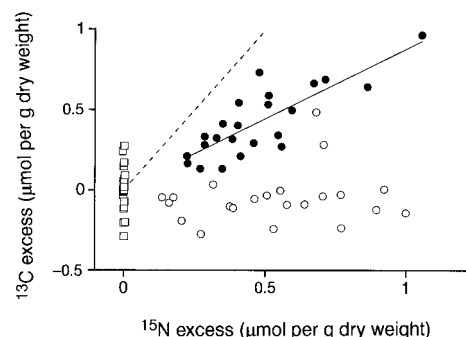


Figure 2 The relationship between excess ^{13}C and excess ^{15}N in the mor layer into which water, or solutions containing $(\text{U-}^{13}\text{C}_2, ^{15}\text{N})$ glycine or $^{15}\text{NH}_4^+$, was injected. The $^{13}\text{C}:^{15}\text{N}$ ratio of the glycine (2:1) injected (broken line) and the regression relating to glycine-injected plots are shown (unbroken line) (slope = 0.87, $r^2 = 0.67$). Data from all samplings are pooled as there were no major differences between samplings. Symbols as in Fig. 1; $n = 24$.

three samplings of the mor layer was 0.86, corresponding to 27% recovery of added glycine carbon (Fig. 2). The rapid decline in the $^{13}\text{C}:$ ^{15}N ratio in the soil between 0 h and 1 day is corroborated by data showing that the ^{13}C - CO_2 respiration rates were 14-fold higher at 2 h than at 7 days. These results further strengthen the conclusion that the turnover of glycine is very dynamic, and support the idea that uptake and initial decarboxylation of the glycine is rapid. The magnitude of that process (Fig. 2) indicates that biotic immobilization of glycine-derived nitrogen is at least as important as any possible mechanism of abiotic immobilization. For example, if all of the 64% of glycine recovered in the mor layer was decarboxylated, there should be 32% recovery of the added ^{13}C . This is similar to the value found (27%, see above), which lends support to the idea that essentially all of the immobilization in the mor layer is biotic. In further support of this, release of ^{15}N after fumigation and extraction, and retention of ^{15}N -labelled tracer in the mor layer, were strongly correlated ($r^2 = 0.78$, $P < 0.001$). The biotic material of the sifted soil will include a fraction of ericaceous hair-roots and extramatrical mycelium, and therefore part of the nitrogen immobilized in this compartment is in the process of being transported to plants.

Evidence accumulated over a long period indicates that nitrogen is a key element regulating the production, structure and function of many terrestrial ecosystems¹. Our data indicate, however, that current conceptual models may be overstating the importance of nitrogen mineralization. We have shown that the dominant plant species in these forests, irrespective of their type of mycorrhiza, all compete well for, and use, glycine as a N source. This demonstrates that nitrogen mineralization is not necessary before plants use organic nitrogen, and suggests that a separation between these groups of plants into those that are capable and incapable of using organic nitrogen should be made with care. However, there may be quantitative differences applying to different organic nitrogen sources, in which the apparently exclusive capacity of ectomycorrhizal and ericoid-mycorrhizal plants to excrete proteases may play a key role⁶. Differences in the capacity of these species to respond to variations in nitrogen supply rate are probably also important¹⁷. □

Methods

Field site. The experiment was conducted at Renberget (64° 14' N, 19° 46' E) in northern Sweden, 60 km northwest of Umeå. The area is located in the middle boreal zone in a late-successional *P. sylvestris*/*P. abies* forest of mesic dwarf shrub type, with a field layer dominated by the ericaceous dwarf shrub *V. myrtillus*. Other species present include the dwarf shrub *V. vitis-idaea* and the grass *D. flexuosa*. The total area used for the study was ~150 m × 50 m. The daytime temperature of the mor layer was ~7–8 °C during the study.

Experimental design. Circular plots (0.0464 m²) were chosen so that shoots of both *V. myrtillus* and *D. flexuosa* plants were present within each plot. The site was divided into eight blocks, and within each block each treatment was replicated three times, allowing three samplings with eight repetitions each ($n = 8$).

Treatments. At 10:00 on 16 June 1996, the experiment was started by applying the labelled substances to the test plots, and water to the control plots. 250 ml 1.2 mM ($\text{U-}^{13}\text{C}_2$, 98 atom percentage; ^{15}N , 96–99 atom percentage) glycine or 1.2 mM ^{15}N - NH_4Cl (98 atom percentage) or water was injected into the mor layer (using a 50-ml syringe fitted with a 4-sideport needle) a few cm above the mineral soil. We made no attempts to stop solutions flowing either down into the mineral soil or laterally in the mor layer.

Sampling. The first step in sampling was cutting and recovering the above-ground vegetation on the plots. Thereafter, below-ground ^{13}C - CO_2 respiration was measured using a cuvette technique¹⁸. These measurements were taken within 10 min, on each of six plots simultaneously. The below-ground compartment was sampled by cutting out the treated mor layer soil core. Samples were put into plastic bags and transported on ice to the laboratory. Above-ground structures of *V. myrtillus* and *D. flexuosa* were sorted so that only fresh young leaves were taken from both species.

Roots of *V. myrtillus*, the two tree species (roots of which were not separated by species) and *D. flexuosa* were carefully picked from the soil core, thoroughly rinsed under tap water to remove soil material, and immersed thrice in 0.5 mM CaCl_2 to remove tracer adsorbed on root surfaces. Then, to obtain soil samples, the remaining cores were sifted through separate sieves (5-mm mesh) for each treatment. Samples were harvested (and stored at –22 °C) ~6 h (2 h for above-ground vegetation), 1 day and 7 days after application of the treatment solutions.

Analyses. Two sets of plant samples were prepared for analysis of ^{13}C and ^{15}N . The first set of plant samples and root-free soil samples were dried (70 °C, 24 h) and milled in a ball mill to a fine powder. The second set of plant samples was prepared by extraction of fresh and frozen roots in 10 mM phosphate buffer (pH 8.0), using 0.5 ml to 200 mg (fresh mass) plant material to obtain the soluble nitrogen fraction. The first set was used directly for isotopic analysis. The solutions of the second set were evaporated under reduced pressure to a final volume of ~50 µl, then transferred to tin capsules with Chromosorb absorbent. The extracts contained ~2–5 mg carbon and 0.02–0.05 mg nitrogen. The solutions were then dried (70 °C) before closing the capsules. Mass-spectrometric analysis was performed on a CF-IRMS¹⁹ (continuous flow isotope ratio mass spectrometer). Microbial nitrogen levels were determined in moist, root-free soil samples after one day of chloroform fumigation and extraction^{20,21}. The atom percentage of nitrogen released from microorganisms was measured after 48 h diffusion²² of distilled nitrogen from Kjeldahl extracts onto acidified Al_2O_3 in tin capsules. The diffusion was done in 12-ml gas-tight vials. Corrections for background nitrogen were made using a calculated blank²³, by comparing diffused and non-diffused isotope standards treated as the original samples. Concentrations of amino acids and inorganic nitrogen species in soil were determined in soil extracts (Table 1) by HPLC.

Calculations. Data on nitrogen and carbon content, and atom percentage ^{15}N and atom percentage ^{13}C in excess of the natural abundances of ^{15}N and ^{13}C in control plots were used to calculate mol ^{15}N excess and mol ^{13}C excess for each sample from plots treated with glycine or NH_4^+ .

Received 9 October 1997; accepted 5 February 1998.

1. Tamm, C.-O. *Nitrogen in Terrestrial Ecosystems* (Springer, Berlin, 1991).
2. Vitousek, P. M. & Howarth, R. W. Nitrogen limitation on land and in sea. How can it occur? *Biogeochemistry* **13**, 87–115 (1991).
3. Melin, E. & Nilsson, H. Transfer of labelled nitrogen from glutamic acid to pine seedlings through the mycelium of *Boletus variegatus* (Sw.) Fr. *Nature* **171**, 134 (1953).
4. Chapin, F. S. III, Moilanen, L. & Kielland, K. Preferential use of organic nitrogen for growth by a non-mycorrhizal arctic sedge. *Nature* **361**, 150–153 (1993).
5. Abuzinadah, R. A. & Read, D. J. The role of proteins in the nitrogen nutrition of ectomycorrhizal plants. III. *New Phytol.* **103**, 507–514 (1986).
6. Read, D. J. Mycorrhizas in ecosystems. *Experientia* **47**, 376–391 (1991).
7. Kaye, J. P. & Hart, S. C. Competition for nitrogen between plants and soil microorganisms. *TREE* **12**, 139–143 (1997).
8. Stark, J. M. & Hart, S. C. High rates of nitrification and nitrate turnover in undisturbed coniferous forest. *Nature* **385**, 61–64 (1997).
9. Eviner, V. T. & Chapin, F. S. III. Plant–microbial interactions. *Nature* **385**, 26–27 (1997).
10. Schimel, J. P. & Chapin, F. S. III. Tundra plant uptake of amino acid and NH_4^+ nitrogen *in situ*: plants compete well for amino acid N. *Ecology* **77**, 2142–2147 (1996).
11. Högborg, P. ^{15}N natural abundance in soil–plant systems. *New Phytol.* **137**, 179–203 (1997).
12. Hultén, E. & Fries, M. *Atlas of North European Vascular Plants North of the Tropic of Cancer Vols 1–III* (Koeltz Scientific Books, Königstein, 1986).
13. Zar, J. H. *Biostatistical Analysis*, 2nd edn (Prentice-Hall, Englewood Cliffs, 1984).
14. Oliver, D. J. The glycine decarboxylase complex from plant mitochondria. *Annu. Rev. Plant Physiol. Mol. Biol.* **45**, 323–337 (1994).
15. Ehleringer, J. R., Field, C. B., Lin, Z. & Kuo, C. Y. Leaf carbon isotope and mineral composition in subtropical plants along an irradiance cline. *Oecologia* **70**, 520–526 (1986).
16. Ekblad, A. & Näsaholm, T. Determination of chitin in fungi and mycorrhizal roots by an improved HPLC analysis of glucosamine. *Plant Soil* **178**, 29–35 (1996).
17. Chapin, F. S. III. The mineral nutrition of wild plants. *Annu. Rev. Ecol. Syst.* **11**, 23–260 (1980).
18. Högborg, P. & Ekblad, A. Substrate-induced respiration measured *in situ* in a C_3 -plant ecosystem using additions of C_4 -sucrose. *Soil Biol. Biochem.* **28**, 1131–1138 (1996).
19. Barrie, A. & Lemley, M. Automated N-15/C-13 analysis of biological materials. *Int. Lab. Techniques* **19**, 82–91 (1989).
20. Brookes, P. C., Landman, A., Pruden, G. & Jenkinson, D. S. Chloroform fumigation and the release of soil nitrogen: a rapid direct extraction method to measure microbial biomass nitrogen in soil. *Soil Biol. Biochem.* **17**, 837–842 (1985).
21. Jenkinson, D. S. in *Advances in Nitrogen Cycling in Agricultural Systems* (ed. Wilson, J. R.) 368–386 (CAB International, Wallingford, 1988).
22. Risgaard-Petersen, N. & Rysgaard, S. in *Methods in Applied Soil Microbiology and Biochemistry* (eds Alef, K. & Nannipieri, P.) 287–295 (Academic, London, 1995).
23. Stark, J. M. & Hart, S. C. Diffusion technique for preparing salt solutions, Kjeldahl digests, and persulfate digests for nitrogen-15 analysis. *Soil Sci. Soc. Am. J.* **60**, 1846–1855 (1996).

Acknowledgements. We thank M. Zetherström and G. Nyberg for assistance in the field and laboratory, and H. Wallmark for conducting the isotope analyses. This work was supported by grants from SJFR (to T.N.) and from NFR, SJFR and CEC (project CANIF) (to P.H.).

Correspondence and requests for materials should be addressed to T.N. (e-mail: Torgny.Nasholm@genfys.slu.se)

A molecular timescale for vertebrate evolution

Sudhir Kumar & S. Blair Hedges

Department of Biology and Institute of Molecular Evolutionary Genetics,
208 Mueller Laboratory, Pennsylvania State University, University Park,
Pennsylvania 16802, USA

A timescale is necessary for estimating rates of molecular and morphological change in organisms and for interpreting patterns of macroevolution and biogeography^{1–9}. Traditionally, these times have been obtained from the fossil record, where the earliest representatives of two lineages establish a minimum time of divergence of these lineages¹⁰. The clock-like accumulation of sequence differences in some genes provides an alternative method¹¹ by which the mean divergence time can be estimated. Estimates from single genes may have large statistical errors, but multiple genes can be studied to obtain a more reliable estimate of divergence time^{1,12–13}. However, until recently, the number of genes available for estimation of divergence time has been limited. Here we present divergence-time estimates for mammalian orders and major lineages of vertebrates, from an analysis of 658 nuclear genes. The molecular times agree with most early (Palaeozoic) and late (Cenozoic) fossil-based times, but indicate major gaps in the Mesozoic fossil record. At least five lineages of placental mammals arose more than 100 million years ago, and most of the modern orders seem to have diversified before the Cretaceous/Tertiary extinction of the dinosaurs.

Molecular clocks are first calibrated with a known time of divergence and then used to estimate divergence times of other species. The divergence of birds and mammals provides a reliable calibration point with which to anchor molecular clocks. The earliest ancestors of mammals (synapsids) and birds (diapsids) are lizard-like and first appear in the Carboniferous period, at ~310 million years (Myr) ago^{10,14} (Fig. 1a). The fact that the fossil record^{10,14} documents a morphological transition from lobe-finned fishes to stem tetrapods at 370–360 Myr ago, and records the appearance of stem amphibians at 338 Myr ago, indicates that the time of the diapsid–synapsid split (within amniotes) is unlikely to be a considerable underestimate. Alternatively, multiple calibration points based on the mammalian fossil record may be used, but this might result in substantial underestimates of divergence time^{12,15}.

We used 658 genes, representing 207 vertebrate species, to estimate divergence times by two methods (see also Fig. 1b; Methods; Supplementary Information). Taxonomic biases in the sequence databases resulted in a predominance of mammalian sequences studied. For estimates derived from large numbers of genes, distributions of divergence-time estimates are approximately normal (Fig. 2a–i; Methods). These distributions show considerable dispersion around the peak, as reflected in high coefficients of variation. On average, standard errors of ~10% were obtained with 10 genes, 5% with 50 genes, and 3% with 100 genes. Multigene time estimates obtained without rate-constancy tests were nearly identical to those obtained with stringent rate testing (Fig. 2j–l). This indicates that there are probably no underlying directional biases in the data.

Molecular times for the origin of the major lineages of vertebrates in the Palaeozoic and early Mesozoic eras are similar to those that are based on the fossil record¹⁴ (Fig. 3). The molecular time estimate for the marsupial–placental split, 173 Myr ago, corresponds well with the fossil-based estimate (178–143 Myr ago)¹⁶. The bird–crocodilian divergence is slightly younger than the earliest fossils suggest¹⁰, at 240 Myr ago, but this difference is less than one standard error. The molecular estimate of the lissamphibian–

amniote divergence at 360 Myr ago also agrees with the fossil-based estimate^{10,14}. Fewer genes are available to time the earliest divergences among vertebrates, but molecular times (of 564 and 528 Myr ago) are consistent with the Late Cambrian fossil record for the first appearance of vertebrates (at 514 Myr ago)¹⁴.

A striking pattern revealed by our molecular divergence times is the Cretaceous origin of all modern orders of mammals examined (Fig. 3). Earlier molecular¹² and fossil¹⁵ studies found evidence that at least some mammalian divergences occurred in the Cretaceous, leaving open the possibility of a gradual diversification of orders into the early Cenozoic. Molecular times now indicate that at least five major lineages of placental mammals (Edentata, Hystricognathi, Sciurognathi, Paenungulata, Archonta + Ferungulata) may have arisen in the Early Cretaceous, >100 Myr ago, and that most mammalian orders were involved in a Cretaceous radiation that predated the Cretaceous/Tertiary extinction of the dinosaurs (Fig. 3). The origin of most mammalian orders seems not to be tied to the filling of niches left vacant by dinosaurs, but is more likely to be related to events in Earth history¹². Similarly, a mid-Cretaceous divergence was obtained for the bird orders Anseriformes and Galliformes^{12,17}.

Multigene divergence times within several orders of mammals compare closely with fossil-based estimates. For example, molecular divergence times from humans to chimpanzees, gorillas, gibbons, and Old World monkeys are close to currently accepted dates from the fossil record^{4,10,18}. The orangutan molecular divergence

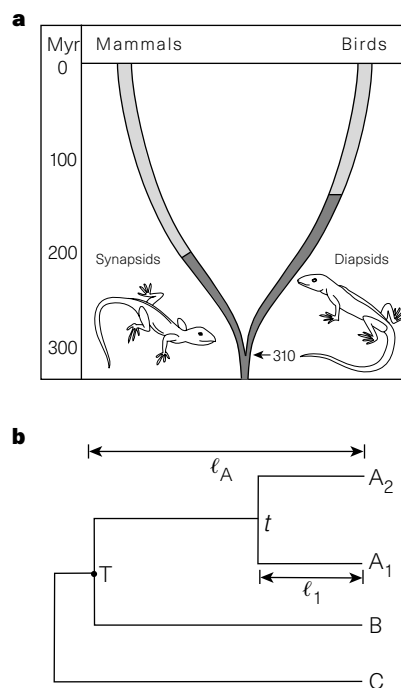


Figure 1 Estimation of divergence times. **a**, Calibration. The arrow marks the first appearance of synapsids (ancestors of mammals) and diapsids (ancestors of birds) in the fossil record at 310 Myr ago. Reconstructions of an early synapsid (*Varanosaurus*) and stem diapsid (*Hylonomus*) are shown. The dark shading represents the reptilian portion and the lighter shading represents the avian and mammalian portion of the phylogeny. **b**, Two methods for dating the unknown divergence time (t) between A_1 and A_2 when A and B diverged at calibration time T (Myr ago). In the average distance method, $t = d_{12}/(2r)$, where $r = (d_{1B} + d_{2B})/(4T)$ is the rate of change for lineages A and B, and d_{ij} is the number of substitutions per site between sequences i and j . In the lineage-specific method, $t_{12} = d_{12}/(2r_A)$ (where $r_A = \ell_A/T$); alternatively, t is based on the length of one of the two lineages; that is, $t = \ell_1/r$ or $t = \ell_1/r_A$, where ℓ_A and ℓ_1 are estimated by the ordinary least squares method (C = outgroup).

time (8.2 Myr ago) corresponds with the age of the unique skull of the fossil hominoid *Sivapithecus*¹⁹, which is usually placed on the orangutan lineage, but is about 4 Myr younger than the earliest teeth and jaw fragments assigned to *Sivapithecus*¹⁹. The *Sivapithecus*–orangutan association itself has been questioned^{20,21}. Molecular time estimates among cetartiodactyls (whales and artiodactyls) and for the catarrhine–platyrrhine and feliform–caniform divergences are close to, or only slightly older than, fossil-based estimates.

In contrast, molecular divergence times among sciurognath rodents (Fig. 3) are roughly four times older than their fossil-based estimates²², as was found previously^{1,7,23}. Because these times were estimated from many genes (343) and did not change when a lineage-specific method (Fig. 1b) was used (e.g., a divergence time of 41 Myr ago was obtained for the mouse–rat divergence), the difference cannot be attributed to stochastic error or increased rate of substitution in rodents. Furthermore, increased stringency of the rate-constancy test resulted in similar time estimates (Fig. 2j–l).

Overall, fossil-based and molecular times are in relatively close agreement (Fig. 4a), except for the origin of placental orders and the early history of rodents. The average difference between molecular and fossil-based dates for Mesozoic comparisons is large (30%) (Fig. 4b). An interpretation of this gap in the Mesozoic fossil record is that molecular times are overestimates. However, this is unlikely as many earlier (Palaeozoic) and later (Cenozoic) fossil and molecular dates show close agreement, especially those dates involving taxa with well documented fossil records and where transitional

forms are known. Recent findings of 85-Myr-old placental fossils from Central Asia^{15–16} are also evidence for the presence of mammalian fossils in this Mesozoic gap.

Our molecular timescale for vertebrate evolution will be useful in calibrating local molecular clocks and in estimating intraordinal divergence times more reliably, especially in groups with poor fossil records. Molecular times also provide an independent measure of the tempo and mode of morphological change. For example, the sudden appearance (in the Early Tertiary fossil record) of mammalian and avian orders, which show large morphological differences, has been taken to imply rapid rates of morphological change at that time^{14,24}. Now, the possibility of 20–70 Myr of prior evolutionary history relaxes that assumption and suggests a greater role for Earth history in the evolution of terrestrial vertebrates^{12,25}. An accurate knowledge of divergence times can help to direct the search for ‘missing’ fossils and test hypotheses of macroevolution. □

Methods

Sequence retrieval and tests of molecular clocks. Amino-acid sequences of nuclear genes were obtained from the HOVERGEN²⁶ database (Genbank Release 97) and all 5,050 gene families were manually examined. Alignments were retrieved whenever data were available to time at least one divergence. Genes under strong positive selection (for example, major histocompatibility complex genes) and sequences with ambiguous orthologies and extensive alignment gaps were excluded. Gene phylogenies were scrutinized further and genes (or sequences) showing extensive rate variation among lineages

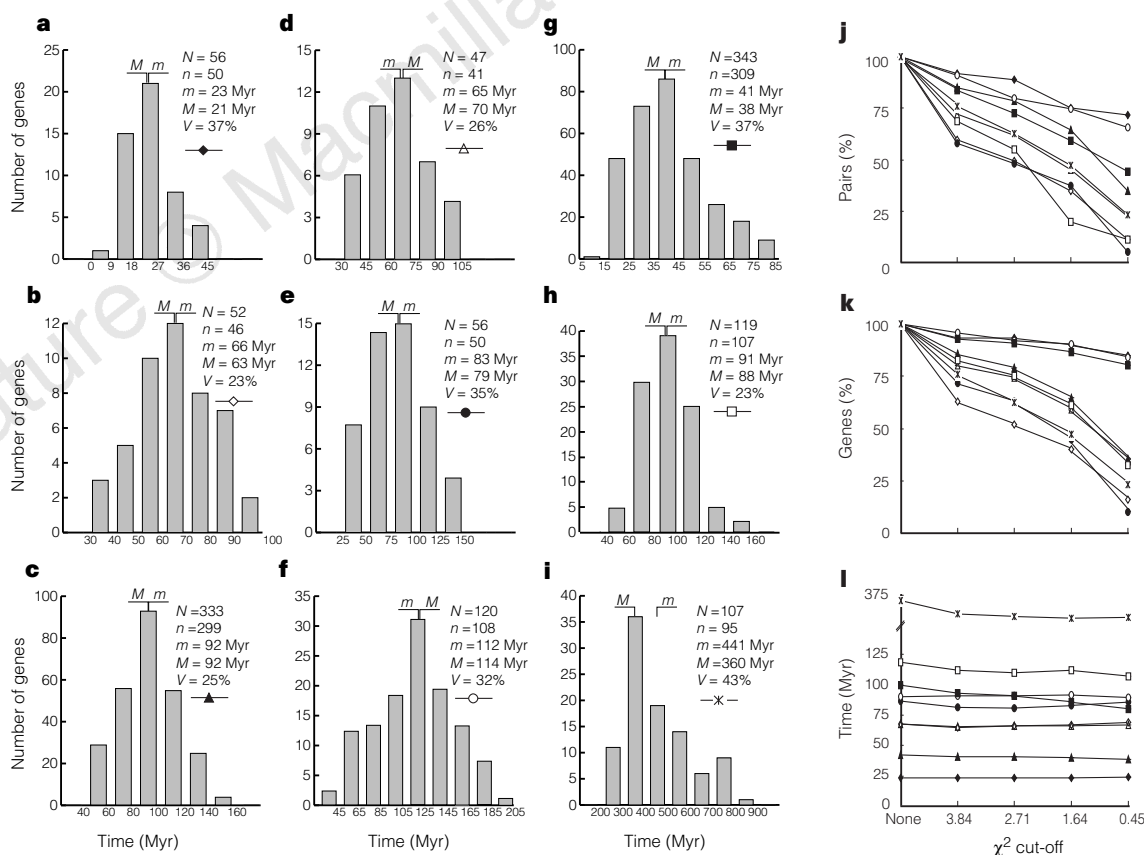


Figure 2 Histograms (a–i) of distributions of single-gene divergence times for nine multigene time estimates, and graphs (j–l) of the effects of increased stringency of the rate-constancy test (corresponding to areas of 5% (recommended), 10%, 20%, and 50%, of the χ^2 rejection curve) for the same divergences. Divergence of: **a**, Hominoidea and Cercopithecoidea; **b**, Muridae and Cricetidae; **c**, Archonta and Ferungulata; **d**, Ruminantia and Suidae; **e**, Carnivora+Perissodactyla and Cetartiodactyla; **f**, Rodentia and (Archonta + Ferungulata+

Paenungulata); **g**, mouse and rat; **h**, Primates and Lagomorpha; and **i**, Amphibia and Amniota. **j**, Percentage of pairwise comparisons not rejected. **k**, Percentage of genes not rejected. **l**, Time estimates for divergences a–i. Symbols for histograms: M, mode; m, mean; N, total number of genes; n, number of genes used after removal of outliers; V, coefficient of variation. The locations of m and M are shown. Myr, millions of years ago.

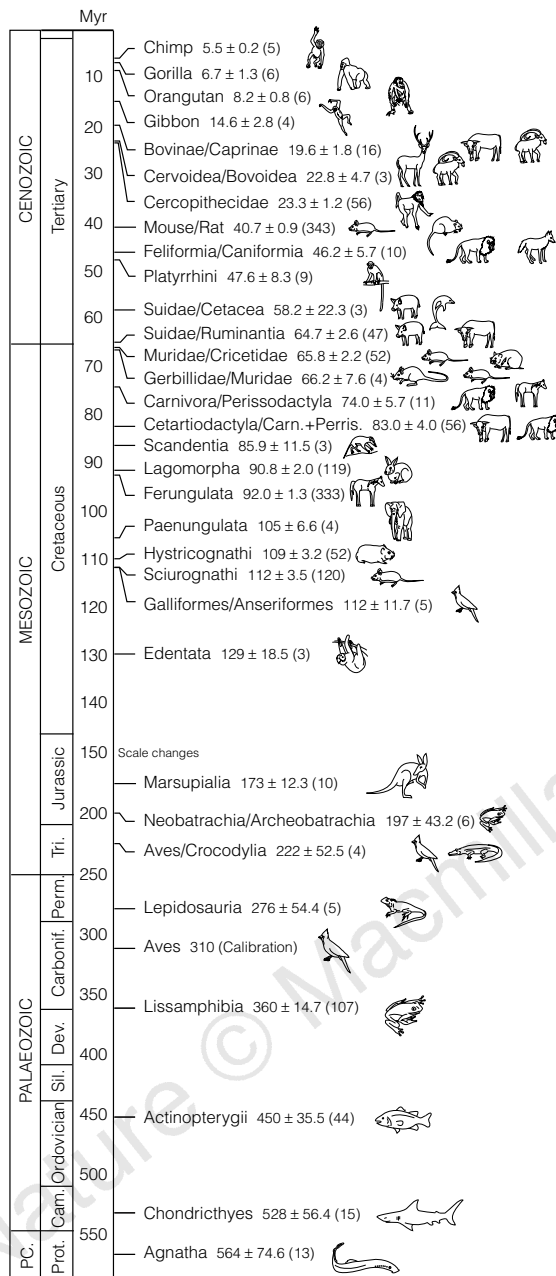


Figure 3 A molecular timescale for vertebrate evolution. All times indicate Myr separating humans (or the largest sister group containing humans) and the group shown, except when the comparative groups are separated by a slash (/). Time estimates are shown with ± 1 s.e.m. and the number of genes used is given in parentheses. Three groups of mammalian orders are: Archonta (Primates, Scandentia, Dermoptera, Chiroptera, and Lagomorpha), Ferungulata (Carnivora, Cetartiodactyla, and Perissodactyla), and Paenungulata (Hyracoidea, Proboscidea, Sirenia). Cam, Cambrian; Carbonif, Carboniferous; Dev, Devonian; PC, Precambrian; Perm, Permian; Prot, Proterozoic; Sil, Silurian; Tri, Triassic.

were removed. The 1DN test²⁷ was conducted for all relevant triplets of sequences, and pairs in which rate constancy was rejected in any of these comparisons were excluded (22% of pairs out of 7,943 pairs were rejected).

Estimating evolutionary rates for protein-clock calibration. Gene-specific evolutionary rates were estimated by $r_{B-M} = d_{avg}/(2 \times 310)$ substitutions per site per Myr, where d_{avg} is the average pairwise sequence divergence between bird (B) (*Gallus gallus* (Gga)) and mammal (M) (*Homo sapiens* (Hsa), *Mus musculus* (Mmu), and/or *Rattus norvegicus* (Rno)) sequences. In the absence of bird sequences, we used a mammalian calibration, based indirectly on the bird-mammal calibration, where the divergence time of rodents and primates (or artiodactyls)¹² of 110 Myr ago was obtained from an analysis of 108 genes; d_{avg} was computed by comparing Mmu, Rno, and Hsa (or *Bos taurus*) sequences. The divergence times estimated using the poisson correction distance (presented here) were similar to those estimated using the gamma distance (shape parameter = 2) (ref. 28).

Estimating average divergence time from multiple genes and species. For each gene, the divergence time between two groups was estimated by taking the average of divergence times from all constant-rate species pairs belonging to those groups (Fig. 1b). This procedure was repeated for every gene, and an average multigene time estimate was calculated. Whenever five or more genes were present, the upper and lower 5% of these estimates were excluded (before averaging) to minimize the influence of outliers (at least the highest and lowest time estimates were excluded). Multigene distributions for the amphibian-amniote (Fig. 2i) and actinopterygian-tetrapod divergences included a higher number of early time estimates. As gene orthology was more difficult to determine for those basal groups, we interpreted such unusually high time estimates as representing paralogous comparisons. Therefore, we used the mode to measure the central value because the mean is sensitive to such outliers. We discarded all estimates that were based on only one or two genes.

Received 14 November 1997; accepted 2 February 1998.

- Wilson, A. C., Carlson, S. S. & White, T. J. Biochemical evolution. *Annu. Rev. Biochem.* **46**, 573–639 (1977).
- Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
- Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356**, 121–125 (1992).
- Martin, R. D. Primate origins: plugging the gaps. *Nature* **363**, 223–234 (1993).
- Avise, J. C. *Molecular Markers, Natural History and Evolution* (Chapman & Hall, New York, 1994).
- Hallam, A. *An Outline of Phanerozoic Biogeography* (Oxford Univ. Press, New York, 1994).
- Eastal, S., Collet, C. & Betty, D. *The Mammalian Molecular Clock* (R. G. Landes, Austin, TX, 1995).
- Gerhart, J. & Kirschner, M. *Cells, Embryos, and Evolution* (Blackwell Scientific, Malden, Massachusetts, 1997).
- Gee, H. *Before the Backbone* (Chapman & Hall, New York, NY, 1996).
- Benton, M. J. *The Fossil Record 2* (Chapman & Hall, London, 1993).
- Zuckerkandl, E. On the molecular evolutionary clock. *J. Mol. Evol.* **26**, 34–46 (1987).
- Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229 (1996).
- Takezaki, N., Rzhetsky, A. & Nei, M. Phylogenetic test of the molecular clock and linearized tree. *Mol. Biol. Evol.* **12**, 823–833 (1995).
- Benton, M. J. *Vertebrate Paleontology* (Chapman & Hall, New York, 1997).
- Archibald, J. D. Fossil evidence for a late Cretaceous origin of "hoofed" mammals. *Science* **272**, 1150–1153 (1996).
- Kielan-Jaworowska, Z. Interrelationships of Mesozoic mammals. *Hist. Biol.* **6**, 185–202 (1992).
- Cooper, A. & Penny, D. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. *Science* **275**, 1109–1113 (1997).
- Kay, R. F., Ross, C. & Williams, B. A. Anthropoid origins. *Science* **275**, 797–804 (1997).
- Ward, S. in *Function, Phylogeny, and Fossils* (eds Begun, D. R., Ward, C. V. & Rose, M. D.) 269–290 (Plenum, New York, 1997).
- Pilbeam, D. Genetic and morphological records of the Hominoidea and hominid origins: a synthesis. *Mol. Phylog. Evol.* **5**, 155–168 (1996).
- Benefit, B. R. & McCrossin, M. L. Earliest known Old World monkey skull. *Nature* **388**, 368–371 (1997).
- Jaeger, J.-J., Tong, H. & Denys, C. The age of the *Mus-Rattus* divergence: paleontological data compared with the molecular clock. *C. R. Acad. Sci. Paris* **302**, 917–922 (1986).
- Parker, P. H. *An Improved Estimate of the Mouse-Rat Divergence Time and Rates of Amino Acid Substitution in Mammals and Birds* Thesis, Pennsylvania State Univ. (1996).
- Carroll, R. L. *Vertebrate Paleontology and Evolution* (W. H. Freeman and Co., New York, 1988).

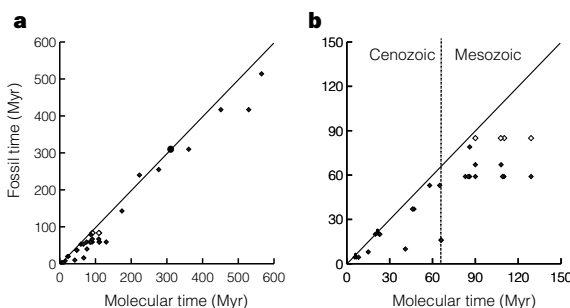


Figure 4 Comparison of fossil-based and molecular estimates of divergence time in vertebrates. **a**, Plot of all values (correlation coefficient = 99%). The solid line indicates a 1:1 relationship; the closed circle represents the calibration point. **b**, Close-up of the region from 0 to 150 Myr ago. Open diamonds show fossil dates based on the 85-Myr placental fossils from Asia^{15,16}, under the assumption that some represent stem (or basal) ferungulates and archontans. Fossil-based estimates^{10,14,16,18,20,22,29} were calculated using a method described elsewhere³⁰. Myr, millions of years ago.

25. Springer, M. S. *et al.* Endemic African mammals shake the phylogenetic tree. *Nature* **388**, 61–63 (1997).
26. Duret, L., Mouchiroud, D. & Gouy, M. HOVERGEN: a database of homologous vertebrate genes. *Nucleic Acids Res.* **22**, 2360–2365 (1994).
27. Tajima, F. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599–607 (1993).
28. Kumar, S., Tamura, K. & Nei, M. MEGA: Molecular Evolutionary Genetic Analysis (Pennsylvania State Univ., 1993).
29. Gheerbrant, E., Sudre, J. & Cappetta, H. A Paleocene proboscidean from Morocco. *Nature* **383**, 68–70 (1996).
30. Benton, M. J. Phylogeny of the major tetrapod groups: morphological data and divergence dates. *J. Mol. Evol.* **30**, 409–424 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We thank L. Poling, A. Beausang, and R. Padmanabhan for assistance with sequence data retrieval; A. Beausang for artwork; A. G. Clark, C. A. Hass, I. Jakobsen, M. Nei, C. R. Rao, and A. Walker for comments and discussion; and L. Duret for instructions on use of the HOVERGEN database. This work was supported in part by grants to M. Nei (NIH and NSF) and S.B.H. (NSF).

Correspondence and requests for materials should be addressed to S.B.H. (e-mail: sbh1@psu.edu).

The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster

Nina M. Brooke*, Jordi Garcia-Fernàndez† & Peter W. H. Holland*

* School of Animal and Microbial Sciences, University of Reading, Whiteknights, PO Box 228, Reading RG6 6AJ, UK

† Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, 08028 Barcelona, Spain

Genes of the Hox cluster are restricted to the animal kingdom and play a central role in axial patterning in divergent animal phyla¹. Despite its evolutionary and developmental significance, the origin of the Hox gene cluster is obscure. The consensus is that a primordial Hox cluster arose by tandem gene duplication close to animal origins^{2–5}. Several homeobox genes with high sequence identity to Hox genes are found outside the Hox cluster and are known as 'dispersed' Hox-like genes; these genes may have been transposed away from an expanding cluster⁶. Here we show that three of these dispersed homeobox genes form a novel gene cluster in the cephalochordate amphioxus. We argue that this 'ParaHox' gene cluster is an ancient paralogue (evolutionary sister) of the Hox gene cluster; the two gene clusters arose by duplication of a

ProtoHox gene cluster. Furthermore, we show that amphioxus ParaHox genes have co-linear developmental expression patterns in anterior, middle and posterior tissues. We propose that the origin of distinct Hox and ParaHox genes by gene-cluster duplication facilitated an increase in body complexity during the Cambrian explosion.

Homeodomain sequence comparisons reveal that at least five classes of homeobox genes are as closely related to Hox genes as many of the latter are to each other⁶. These are the Evx, Mox, Cdx (or cad), Xlox, and Gsx homeobox classes (we term a class defined by mouse *Gsh-1* and *Gsh-2* as Gsx). The two mammalian Evx genes are each linked to the 5' end of Hox gene clusters⁶, and a cnidarian Evx-like gene is linked to a Hox-like gene⁷, indicating that the close sequence relationship between Evx and Hox genes reflects tandem duplication. Mox genes may represent a similar case because the mouse *Mox-1* gene maps to chromosome 11, close to the Hox cluster⁸. The Cdx, Xlox and Gsx gene families are more problematic.

To investigate the evolutionary origins of Cdx, Xlox and Gsx genes, we elected to clone representatives of each gene family from amphioxus. This is because homeobox gene families in this animal are not complicated by either excessive duplication (as in vertebrates⁹) or divergence and rearrangement (as in *Drosophila* or nematode)^{6,10}. Using primers directed to Hox class homeoboxes and amphioxus genomic DNA as template, amplification by polymerase chain reaction (PCR) yielded partial clones of Cdx and Xlox class homeoboxes. A fragment of amphioxus Gsx was also cloned by PCR, using primers designed from the two mammalian gene family members *Gsh-1* and *Gsh-2*. To determine the complete homeobox sequence of each gene, we isolated longer clones from amphioxus genomic libraries: only single members of each class were obtained, which we named *AmphiCdx*, *AmphiXlox* and *AmphiGsx*. Their encoded homeodomains resemble those of the *Drosophila* or vertebrate homologues (Fig. 1).

Analysis of genomic clones revealed that amphioxus Xlox and Cdx class homeoboxes were unexpectedly contained within a single bacteriophage clone. Mapping indicated that the homeoboxes were separated by just 7.5 kilobases (kb). Furthermore, using genomic walking we found that these two homeobox genes are physically linked to the *AmphiGsx* gene. The Gsx and Xlox homeoboxes are separated by just 25 kb (Fig. 2). We designate this tight cluster of three genes the ParaHox gene cluster.

The finding that amphioxus Gsx, Xlox and Cdx class genes form a novel homeobox cluster challenges the idea that these homeobox gene classes are 'dispersed' Hox genes. To reconcile linkage in

AmphiCdx	KDKYRVVSDHQRLEKEFEYSNKYITIKRQVLNELGLSERQVKIWFQNRRAKQKRMA	100%
mCdx-1	-----T-----HYSR--R--SE--AN--T-----E--VN	77%
mCdx-2	-----T-----HFSR--R--SE--AT-----E--IK	78%
mCdx-4	-E-----T-----HC-R--R--SE--VN-----E--MIK	77%
D Cad	-----T-F-----YCTSR--R--SE--QT-S-----E-TSN	72%
Ce pal-1	A--M--Y-----HTSPF--SD--S--STM-S-T--I-----D-RDK	65%
AmphiXlox	NKRTRTAYTRGQLEKEFEFNKYISRPRRIELAAMLNLTERHIKIQNRMRKWKKEQ	100%
mpdx-1	-----A-----L-----V--V-----E	92%
XlHbox8	-----A-----L-----V--V-----E	92%
Htr-A2	-----S-S-F-----D-----V--SS-----ME	85%
AmphiGsx	SRMRRTAFSSTQLELEREFASMYLSRLRIEATFLNLSEKQVKIWFQNRVRKHKKEA	100%
mGsh-1	-K-----T-----Y-----G	93%
mGsh-2	GK-----T-----S-----Y-----G	90%

Figure 1 Homeodomains of amphioxus Cdx, Xlox and Gsx genes aligned to mouse (m), *Drosophila* (D), nematode (Ce), *Xenopus* (Xl) and leech (Htr) homologues. The mouse Cdx2 gene is the probable orthologue of human CD33. Dashes indicate identical amino acids.

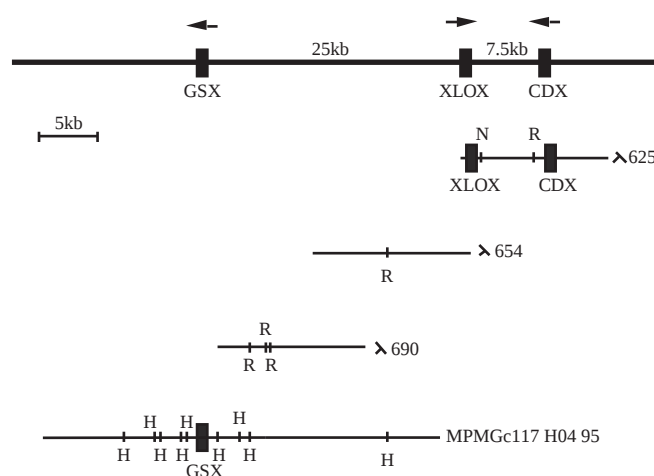


Figure 2 Genomic organization of amphioxus Gsx, Xlox and Cdx genes, showing genomic clones used in walking. Arrows denote transcriptional orientation. R, *EcoRI* site and N, *NotI* site, mapped in bacteriophage clones only; H, *HindIII* sites mapped in cosmid only.

amphioxus with an origin by transposition, *Gsx*, *Xlox* and *Cdx* genes must have been transposed from the evolving Hox gene cluster as a cassette of three adjacent genes. We tested this by molecular phylogenetic analysis of Hox, *Gsx*, *Xlox* and *Cdx* genes (Fig. 3) and found that the *Gsx*, *Xlox* and *Cdx* gene classes do not form a distinct branch on the tree, as would be predicted for genes that arose as neighbours within an early Hox gene cluster. The three gene classes are widely spread in the tree topology. To test the statistical significance of this finding, we estimated likelihood values for the tree shown in Fig. 3 and for 15 modified trees consistent with transposition of a cassette (*Gsx*, *Xlox* and *Cdx* forming a distinct clade basal to any Hox subfamily). A resampling of estimated likelihood method found that the tree shown in Fig. 3 is supported over the modified trees with a bootstrap value of 97.6%.

The tree topology suggests a probable evolutionary origin for Hox

and ParaHox gene clusters. Within the tree, Hox genes divide into four subfamilies of evolutionarily related genes: 'anterior' genes (chordate PG1, PG2; insect *lab*, *pb*), 'group 3' genes (chordate PG3; insect *zen*/PG3), 'middle' genes (chordate PG4 to PG8; insect *Dfd* to *abd-A*) and 'posterior' genes (chordate PG9 to PG13; insect *Abd-B*). These subfamilies may reflect a stage in Hox gene evolution consisting of four linked genes. We find that *Gsx*, *Xlox* and *Cdx* class genes are grouped with different Hox gene subfamilies: the anterior, group 3 and posterior subfamilies, respectively. On the basis of this molecular phylogeny combined with the physical linkage data, we propose that Hox and ParaHox gene clusters arose by duplication of an ancestral 'ProtoHox' gene cluster. This ancestral state probably consisted of four linked genes, although more or less cannot be discounted with confidence. After duplication of the ProtoHox cluster, one set of descendant genes underwent a series of tandem duplications to yield the definitive Hox gene cluster, while the other evolved into the ParaHox gene cluster of *Gsx*, *Xlox* and *Cdx* (Fig. 4). We use the term ParaHox because this gene cluster is a paralogue of the definitive Hox gene cluster.

To investigate the possible functions of the ParaHox gene cluster, we examined the developmental expression patterns of *AmphiCdx*, *AmphiXlox* and *AmphiGsx*. The level of *AmphiGsx* expression is low, and is detected only in the cerebral vesicle (fore-/midbrain homologue¹¹) of 18–20-h embryos (Fig. 5a). *AmphiXlox* is strongly expressed in a stripe of the archenteron wall (presumptive gut) from the neurula stage through to larvae (Fig. 5b). The region of expression is just anterior to the posterior pole at the developmental stage shown. As development proceeds, the body grows at the posterior end, so that *AmphiXlox*-expressing cells become far anterior to the anus in the swimming larvae (data not shown). There is also strong but transient expression in two cells of the

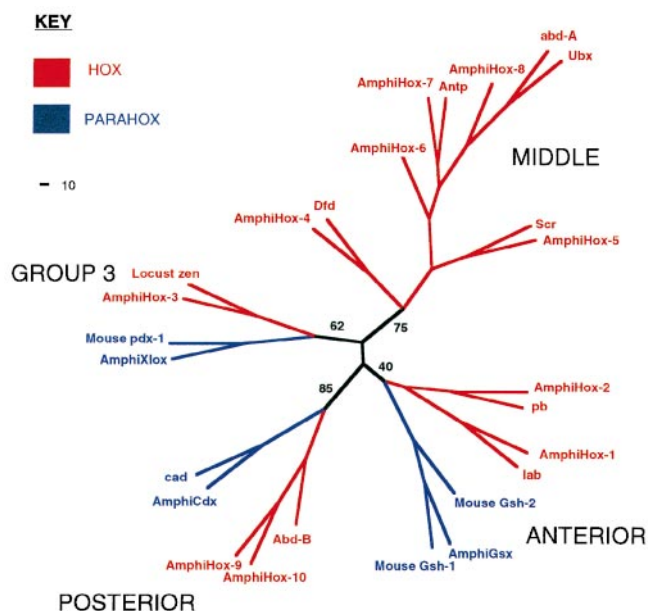


Figure 3 Evolutionary relationships of *Cdx*, *Xlox* and *Gsx* genes to Hox genes, as deduced by neighbour-joining analysis of homeodomains. The tree is unrooted and shows four subfamilies. Figures refer to bootstrap percentages.

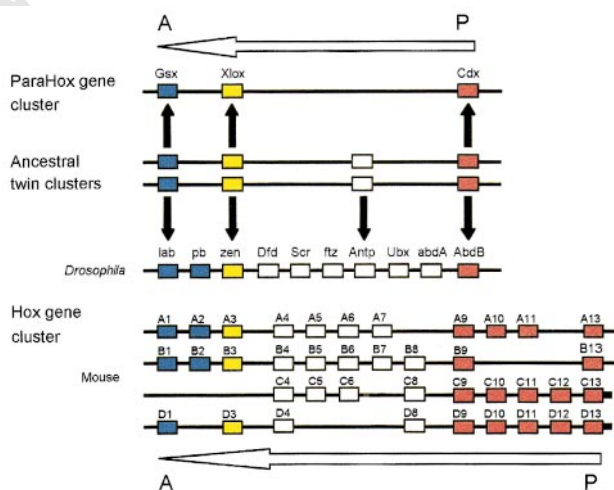


Figure 4 Origin of Hox and ParaHox gene clusters inferred from combining gene linkage and phylogenetic analyses. Hox and ParaHox gene clusters evolved from ancestral twin clusters generated by duplication. Horizontal arrows denote polarity of spatial colinearity (A, anterior; P, posterior).

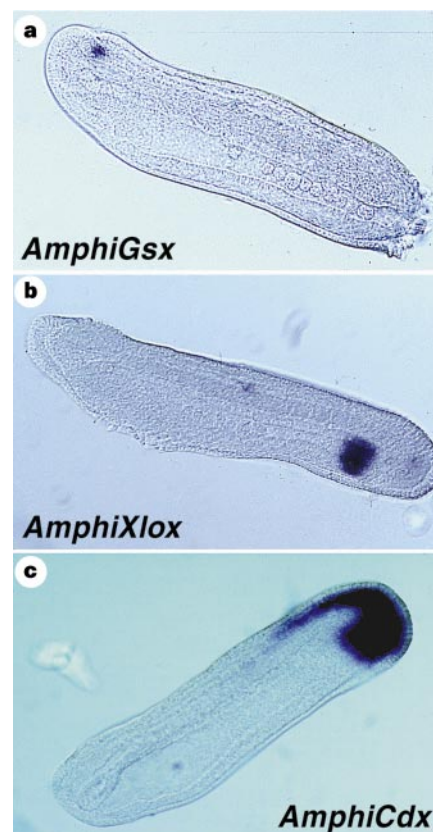


Figure 5 Whole-mount *in situ* hybridization to 20-h amphioxus embryos. Expression of **a**, *AmphiGsx*; **b**, *AmphiXlox*; **c**, *AmphiCdx* genes. Anterior is to the left.

neural tube opposite somite 5, at a site fated to form the first pigment spot. This has faded to a faint signal at the stage shown in Fig. 5b. In early neurulae, *AmphiCdx* is expressed posteriorly in all tissues. As development proceeds, expression persists most strongly in the posterior region of the developing gut, continuous with a gradient of expression in the most posterior part of the neural tube (Fig. 5c). At 20 h development, *AmphiCdx* expression in the gut is just posterior to that of *AmphiXlox* (Fig. 5b, c); by the larval stage, the two expression domains are clearly separated.

Data from other species argue that ParaHox gene expression sites have been conserved in evolution. Like amphioxus, both arthropods and vertebrates express Cdx genes in the hindgut^{12,13}. The multiple Cdx genes of vertebrates have additional roles in other tissues but these are not shared between gene family members and may be derived. Expression of Xlox in central regions of the presumptive gut has been reported in vertebrates^{14,15} and leeches¹⁶ and is closely comparable to *AmphiXlox*. The single mammalian Xlox gene *pdx-1* is expressed in the endodermal precursor of the pancreas and is required for correct development of pancreas and rostral duodenum^{15,17}. Like *AmphiGsx*, two Gsx genes previously described from mouse are expressed in the developing brain. *Gsh-2* is involved in brain development¹⁸ and *Gsh-1* in development of the adenohypophysis¹⁹. This component of the pituitary develops as an ectodermal outpocketing of the oral cavity (Rathke's pouch); the homologous structure is endodermal in hagfish²⁰. Hence, mouse Gsx genes have roles in both brain and anterior-gut development. We cannot be certain whether *AmphiGsx* is also expressed in anterior gut; its expression is at the limit of sensitivity of our methodology and weak expression in a few anterior gut cells would remain undetected.

In summary, the order of genes along the ParaHox gene cluster (Fig. 2) matches the spatial order of gene-expression sites along the body axis (Fig. 4). Hence, ParaHox genes, like definitive Hox genes²¹, display spatial co-linearity, although the details differ. ParaHox genes obey the co-linearity rule in the neural tube and at least two, possibly three, of the genes also obey this rule in the gut. The polarity of spatial co-linearity is identical between Hox and ParaHox genes (anterior genes related to anterior genes and so on), suggesting that both genetic systems inherited spatial co-linearity from the common precursor gene cluster.

Precise dating of Hox/ParaHox origins is not yet possible, but the cluster duplication must predate the divergence of 'higher' triploblast phyla. This is because gene clusters containing definitive Hox genes are present in arthropods, echinoderms, nematodes and chordates^{1,6,9,22}, whereas genes belonging to the three ParaHox classes have been reported in arthropods, annelids, nematodes, echinoderms and chordates^{6,13,15,17,23}. Circumstantial evidence points to retention of the ParaHox gene linkages in mammals, despite further duplication. Thus, the human *PDX1* gene maps to the same chromosomal location as one of the Cdx genes (*CDX3*) at 13q12.1 (ref. 24), and mouse *Gsh-2* maps to the region syntenic to 13q12.1 (chromosome 5 at 89 cM)^{25,26}. The situation in more basal lineages of animals is unclear. Our phylogenetic analyses of cnidarian Cnox genes give ambiguous results, and more physical linkage data from cnidarians are needed to clarify the situation. Current data are consistent with the hypothesis that distinct Hox and ParaHox gene clusters originated on the triploblast stem lineage. The dramatic increase in body-plan diversity associated with the Cambrian explosion probably also occurred on this lineage²⁷ and predominantly affected animals of the triploblast grade. Duplication of a primordial ProtoHox gene cluster, followed by functional recruitment to different tissues, may have facilitated an increase in body complexity during the Cambrian explosion. □

Methods

Genomic cloning. Total DNA was extracted from adult *Branchiostoma floridae* from Old Tampa Bay, Florida, for PCR and library construction. Fragments of

Cdx and Xlox were isolated by PCR using primers SO1 and SO2 (ref. 9). Gsx amplification used SO2 in combination with primer Sogsh (CAGCTCTTG-GARCTNGARCGNG). Genomic clones containing the Cdx and Xlox genes were isolated from a lambda FIXII library⁹. Gsx was isolated from cosmid library MPMGc117 of the RZPD, Berlin (ref. 28). Genomic walking was performed as described⁹. Overlap of phage and cosmid clones was verified by subcloning and sequencing a shared 1-kb *HindIII* fragment.

Phylogenetic analyses. Molecular phylogenetic trees were constructed from homeodomain protein sequences, to avoid possible distortion by codon usage differences, and utilized the neighbour-joining method on a Kimura distance matrix, implemented using PHYLIP3.572c (ref. 29). Topology robustness was assessed by bootstrap resampling. Statistical likelihood of alternative user-defined trees was assessed by a resampling of estimated likelihood method implemented with PROTML in PHYLIP3.572c (ref. 29).

Whole mount *in situ* hybridization. Hybridization to amphioxus embryos was done as described³⁰, except that detection time was increased to 5 days for *AmphiGsx*.

Received 4 December 1997; accepted 24 February 1998.

- Slack, J. M. W., Holland, P. W. H. & Graham, C. F. The zootype and the phylotypic stage. *Nature* **361**, 490–492 (1993).
- Kappen, C., Schughart, K. & Ruddle, F. H. Two steps in the evolution of Antennapedia-class vertebrate homeobox genes. *Proc. Natl Acad. Sci. USA* **86**, 5458–5463 (1989).
- Schubert, F. R., Nüssli-Sturwe, K. & Gruss, P. The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. *Proc. Natl Acad. Sci. USA* **90**, 143–147 (1993).
- Gehring, W. J., Affolter, M. & Bürglin, T. Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487–526 (1994).
- Zhang, J. & Nei, M. Evolution of Antennapedia-class homeobox genes. *Genetics* **142**, 295–303 (1996).
- Bürglin, T. in *A Guidebook for Homeobox Genes* (ed. Duboule, D.) 25–71 (Oxford University Press, Oxford, 1994).
- Miller, D. J. & Miles, A. Homeobox genes and the zootype. *Nature* **365**, 215–216 (1993).
- Candia, A. F. et al. *Max-1* and *Max-2* define a novel homeobox gene subfamily and are differentially expressed during early mesodermal patterning in mouse embryos. *Development* **116**, 1123–1136 (1992).
- García-Fernández, J. & Holland, P. W. H. Archetypal organization of the amphioxus Hox gene cluster. *Nature* **370**, 563–566 (1994).
- Averof, M., Dawes, R. & Ferrier, D. Diversification of arthropod Hox genes as a paradigm for the evolution of gene functions. *Sem. Cell. Dev. Biol.* **7**, 539–551 (1996).
- Williams, N. A. & Holland, P. W. H. Old head on young shoulders. *Nature* **383**, 490 (1996).
- Duprey, P. et al. A mouse gene homologous to the *Drosophila* gene *caudal* is expressed in epithelial cells from the embryonic intestine. *Genes Dev.* **2**, 1647–1654 (1988).
- Calleja, M., Moreno, E., Pelaz, S. & Morata, G. Visualization of gene expression in living adult *Drosophila*. *Science* **274**, 252–255 (1996).
- Wright, C. V. E., Schnegelsberg, P. & DeRobertis, E. M. *XIHbox 8*—a novel *Xenopus* homeoprotein restricted to a narrow-band of endoderm. *Development* **104**, 787–794 (1988).
- Offield, M. F. et al. *Pdx-1* is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122**, 983–995 (1996).
- Wysocka-Diller, J., Aisemberg, G. O. & Macagno, E. R. A novel homeobox cluster expressed in repeated structures of the midgut. *Dev. Biol.* **171**, 439–447 (1995).
- Jonsson, J., Carlsson, L., Edlund, T. & Edlund, H. Insulin-promoter-factor-1 is required for pancreas development in mice. *Nature* **371**, 606–609 (1994).
- Szucsik, J. C. et al. Altered forebrain and hindbrain development in mice mutant for the *Gsh-2* homeobox gene. *Dev. Biol.* **191**, 230–242 (1997).
- Li, H., Zeidler, P. S., Valerius, M. T., Small, K. & Potter, S. S. *Gsh-1*, an orphan Hox gene, is required for normal pituitary development. *EMBO J.* **15**, 714–724 (1996).
- Gorbman, A. Early development of the hagfish pituitary gland: evidence for the endodermal origin of the adenohypophysis. *Am. Zool.* **23**, 639–654 (1983).
- Gaunt, S. J., Sharpe, P. T. & Duboule, D. Spatially restricted domains of homeogene transcripts in mouse embryos: relation to a segmented body plan. *Development* (suppl.) **104**, 169–180 (1988).
- Popodi, E., Kissinger, J. C., Andrews, M. E. & Raff, R. A. Sea urchin Hox genes: insights into the ancestral Hox cluster. *Mol. Biol. Evol.* **13**, 1078–1086 (1996).
- Irvine, S. Q., Warinner, S. A., Hunter, J. D. & Martindale, M. Q. A survey of homeobox genes in *Chaetopterus variopedatus* and analysis of polychaete homeodomains. *Mol. Phylo. Evol.* **7**, 331–345 (1997).
- Stoffel, M. et al. Localization of human homeodomain transcription factor insulin promoter factor-1 (*Ipfl*) to chromosome band 13q12.1. *Genomics* **28**, 125–126 (1995).
- Kozak, C. A., Goffinet, A. & Stephenson, D. A. Mouse chromosome-5. *Mamm. Genome* **5**, s65–s78 (1994).
- Fiedorek, F. T. & Kay, E. S. Mapping of the insulin promoter factor-1 gene (*Ipfl*) to distal mouse chromosome-5. *Genomics* **28**, 581–584 (1995).
- Conway Morris, S. Why molecular biology needs palaeontology. *Development* (suppl.) 1–13 (1994).
- Lehrach, H. et al. in *Genome Analysis 1: Genetic and Physical Mapping* (eds Davies, K. E. & Tilghman, S. M.) 38–81 (Cold Spring Harbor Laboratory Press, New York, 1990).
- Felsenstein, J. PHYLIP version 3.5c (Department of Genetics, Univ. Washington, Seattle, 1993).
- Holland, P. W. H., Holland, L. Z., Williams, N. A. & Holland, N. D. An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution. *Development* **116**, 653–661 (1992).

Acknowledgements. We thank G. Balavoine and C. V. E. Wright for discussion, S. J. Patton and C. Burgdorf for access to the cosmid library and help with screening, and N. A. Williams and S. A. J. Thompson for advice and assistance. This research was funded by a BBSRC Earmarked Studentship (to N.M.B.) and by the DGICYT (J.G.F.), and facilitated by a grant from the Acciones Integradas of the British Council/Ministerio de Educación y Ciencia.

Correspondence and requests for materials should be addressed to P.W.H.H. (e-mail: p.w.h.holland@reading.ac.uk). The *AmphiGsx*, *AmphiXlox* and *AmphiCdx* sequences are deposited in GenBank under accession numbers AF052463 to AF052465.

Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation

Pierre Billuart*, Thierry Bienvenu*, Nathalie Ronce†, Vincent des Portes*, Marie Claude Vinet*, Ramzi Zemni*, Hugues Roest Crollius‡, Alain Carrié*, Fabien Fauchereau*, Michele Cherry§, Sylvain Briault†, Ben Hamel||, Jean-Pierre Fryns¶, Cherif Beldjord*, Axel Kahn*, Claude Moraine† & Jamel Chelly*

*INSERM U129-ICGM, Faculté de Médecine Cochin, 24 rue du Faubourg Saint-Jacques, 75014 Paris, France

†Centre Hospitalier Universitaire de Tours, Service de Génétique, Hôpital Bretonneau, 2 boulevard Tonnelles, 37044 Tours Cedex, France

‡Max-Planck-Institute for Molecular Genetics, Ihnestrass 73, Berlin-Dahlem, Germany

§Centre Hospitalier Universitaire de Nancy, Laboratoire de Génétique, Rue du Morvan, 54511 Vandœuvre les Nancy Cedex, France

||University Hospital Nijmegen, 417 Department of Human Genetics, Geert Grooteplein 10, 6500 HB Nijmegen, The Netherlands

¶Center for Human Genetics, Clinical Genetics Univ, UZ Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium

Primary or nonspecific X-linked mental retardation (MRX) is a heterogeneous condition in which affected patients do not have any distinctive clinical or biochemical features in common apart from cognitive impairment¹. Although it is present in approximately 0.15–0.3% of males², most of the genetic defects associated with MRX, which may involve more than ten different genes, remain unknown³. Here we report the characterization of a new gene on the long arm of the X-chromosome (position Xq12) and the identification in unrelated individuals of different mutations that are predicted to cause a loss of function. This gene is highly expressed in fetal brain and encodes a protein of relative

molecular mass 91K, named oligophrenin-1, which contains a domain typical of a Rho-GTPase-activating protein (rhoGAP)^{4,5}. By enhancing their GTPase activity, GAP proteins inactivate small Rho and Ras proteins, so inactivation of rhoGAP proteins might cause constitutive activation of their GTPase targets. Such activation is known to affect cell migration and outgrowth of axons and dendrites *in vivo*^{6–8}. Our results demonstrate an association between cognitive impairment and a defect in a signalling pathway that depends on a Ras-like GTPase.

We have previously identified a female patient with an X;12 balanced translocation associated with mild mental retardation⁹. Figure 1 shows the location of the translocation breakpoint on the normal X and artificial chromosomes (YAC from yeast; PAC from phage) and cosmid contigs spanning the breakpoint. The probe STSC16T3, which detects the derivative 12 (der12) junction fragment and localized the breakpoint to a 9-kilobase (kb) *Hind*III fragment (Fig. 1b), was isolated from the cosmid clone 4D2 (Fig. 1a).

To identify the expressed sequences in the vicinity of the breakpoint, we sequenced randomly subcloned *Hind*III fragments and searched for homology in the databases. This approach revealed sequence identities between our sequences and those corresponding to the PAC clone 360E18 (Fig. 1a) generated by the Sanger Centre (Cambridge, UK). The retrieved sequences were then used for computational analysis and compared with nucleotide and protein databases. Some of the potential exon sequences, identified by using GRAIL¹⁰ and FEXH/HEXON¹¹ programs, and their predicted peptides showed significant similarity with the mouse expressed sequence tag (EST) 483210, the human EST 387042 on chromosome 11, and the *Caenorhabditis elegans* open reading frame (ORF) CELT04C95. We confirmed these predictions in reverse transcription–polymerase chain reaction (RT-PCR) experiments using primers located in the potential exons and human fetal brain total RNA as template, and detected the expected RT-PCR fragments (data not shown). We hybridized the RT-PCR products to a northern blot containing polyadenylated RNA (poly(A)⁺ RNA) and detected a 7.5-kb transcript that is highly expressed in human fetal brain tissue (Fig. 2). Together, these results indicated the presence of a candidate gene in

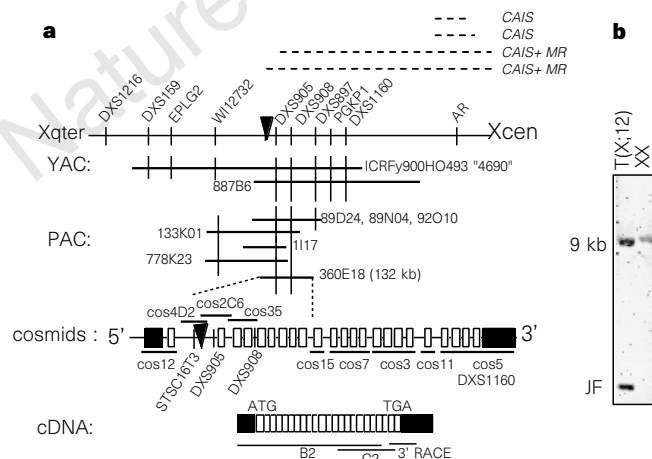


Figure 1 Genomic region containing Xq12 translocation breakpoint and oligophrenin-1 gene. **a**, Physical map of the Xq12 locus and genomic structure of the *oligophrenin-1* gene. The arrowhead indicates the X-chromosome breakpoint. The *oligophrenin-1* gene spans at least 300 kb. Dotted lines represent deletions in patients either with complete androgen-insensitivity syndrome (CAIS) and mental retardation or with CAIS¹³. Mapping data show that deletions in patients with mental retardation extend to the *oligophrenin-1* gene, whereas deletions in patients with only CAIS are limited to the androgen-receptor gene. **b**, Southern blot analysis of *Hind*III-digested genomic DNAs hybridized with STS C16 T3, junction fragment.

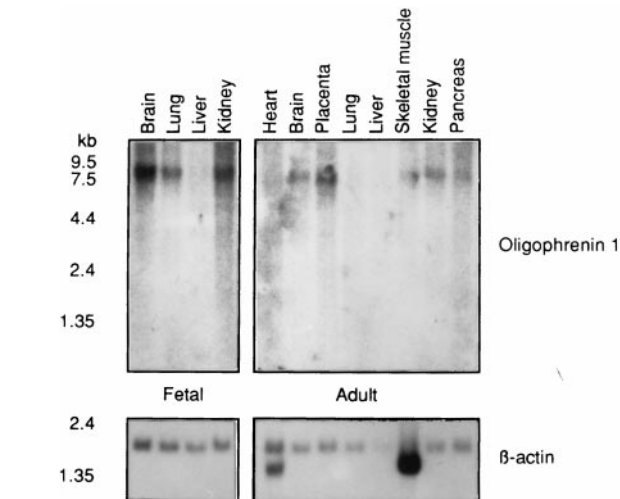


Figure 2 Fetal and adult multiple-tissue northern blots containing poly(A)⁺ RNA hybridized with the C2 cDNA clone from a fetal brain cDNA library. A 7.5-kb transcript was observed after an overnight exposure. The same transcript was detected with RT-PCR products generated with primers derived from the predicted exons. Northern blots were hybridized with an actin probe to check for equal loading of poly(A)⁺ RNA samples.

the vicinity of the translocation breakpoint. We then isolated and sequenced the full-length complementary DNA (see Methods and Fig. 4), and determined the structure of the gene, including the exon-intron boundaries (data not shown).

Physical mapping of the 25 exons and fluorescence *in situ* hybridization (FISH) analysis using the cosmid clone 12 (Fig. 1a) as probe indicated that the translocation breakpoint maps within the second intron. The X;12 translocation therefore resulted in a displacement on the der12 of the first two exons, one of which contains the putative translation initiation codon. To investigate the effect on gene expression and to test whether both alleles were inactive, we did RT-PCR on RNA isolated from Epstein-Barr virus (EBV)-transformed lymphoblastoid cell lines of the patient using primers located in exons 1 and 2, but we failed to amplify a normal gene product on the patient's RNA (Fig. 3a) and demonstrated the absence of expression of this candidate gene in the patient.

To demonstrate the involvement of this gene in X-linked mental retardation, we searched for mutations in four probands from MRX families mapped in genetic intervals that encompass the candidate gene (Table 1). Denaturing gradient-gel electrophoresis¹² analysis (DGGE) and sequencing of PCR products corresponding to all coding exons revealed the presence in the proband IV-3 of the MRX60 family of a one-base-pair deletion corresponding to the nucleotide 1,578 (part of the valine codon 526; Fig. 4a). Co-segregation of this frameshift mutation with the disease was confirmed in the large family by DGGE (Fig. 3b). Also, RT-PCR analysis using RNA isolated from a lymphoblastoid cell line from the proband revealed that this mutation resulted in a barely detectable amount of the isolated transcript (data not shown). This frameshift mutation was not detected in 100 control individuals.

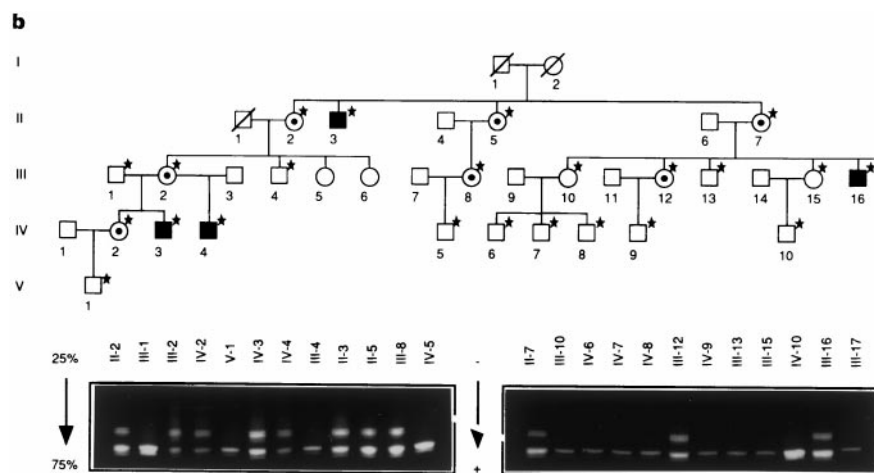
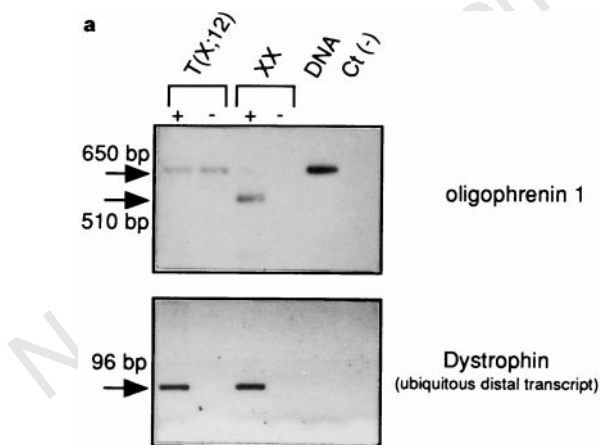


Table 1 Summary of linkage data in affected families

Locus	Family number	Z _{max} at marker	Flanking markers
Xp11.4-q12	MRX60	3.01 at AR	OTC/DXS981
Xp11.3-q12	MRX14	3.31 at DXS255	PFC/DXS135
xp11.21-q21.3	MRX52	3.14 at DXS559	ALAS2/DXS1002
XP11.23-q21.3	F91-09	3.64 at ALAS2	DXS573/DXS990

Genetic intervals identified in these families encompass the *oligophrenin-1* gene. Z_{max} is the maximum lod score at the indicated marker. Flanking markers are the recombinant markers that define the genetic interval.

Our results indicate that loss of function of this gene, which we name *oligophrenin-1*, is responsible for X-linked nonspecific mental retardation; our conclusion is supported by deletions described elsewhere¹³ and summarized in Fig. 1. Relevant mutations were not detected in the 23 coding exons of *oligophrenin-1* in probands from the three other families. Analysis of *oligophrenin-1* transcript by RT-PCR using RNA isolated from lymphoblastoid cells from two probands did not detect significant difference when compared with the expression in lymphoblastoid cells from normal individuals (data not shown). Although our findings cannot exclude the possibility of mutations in the promoter or intronic sequences, a second locus in the pericentromeric region might be involved in MRX.

The open reading frame of *oligophrenin-1* is 2,406 bases and begins with a methionine codon preceded by a consensus sequence for translation initiation (ACCAUGG)¹⁴. It encodes a predicted protein of 802 amino acids (Fig. 4a). Comparison of the predicted protein sequence with other sequences in the databases indicated that oligophrenin 1 contains a rhoGAP domain (Fig. 4b), and its amino-terminal domain is similar to a highly conserved protein, of unknown function, identified in *C. elegans*, mouse and human (Fig. 4c). To determine whether the oligophrenin-1 rhoGAP domain possesses GAP activity and study its specificity, we expressed it as a fusion protein with glutathione S-transferase (GST-OP1). The purified fusion protein was incubated with various recombinant small GTPases preloaded with [γ -³²]GTP and the effect on the rate of GTP hydrolysis was investigated. The purified GSP-OP1 protein stimulated GTP hydrolysis of RhoA, Rac1 and Cdc42Hs (Fig. 5a-c). In contrast, the fusion protein did not stimulate GTP hydrolysis of Rap2 (Fig. 5b, c), a Ras-like GTPase that does not belong to the Rho subfamily. These *in vitro* results indicate that oligophrenin 1 specifically stimulates GTP hydrolysis of members of the Rho subfamily.

The *oligophrenin-1* transcript was most abundant in RNA from fetal brain (Fig. 2) We used *in situ* hybridization to examine the expression of the mouse homologous gene during development. The gene is expressed at low levels in all tissues and at a higher level

Figure 3 Expression of oligophrenin-1 in the patient's cell lines and detection of mutation in the MRX60 family. **a**, Southern blot of RT-PCR products corresponding to *oligophrenin-1* and internal standard transcripts. PCR was done in the presence (+) or absence (-) of reverse transcriptase. The 650-bp fragment is amplified from genomic DNA present in the RNA preparation (intron 1 is 140 bp). DNA, PCR using genomic DNA as template; Ct(-), PCR without template. The expected 510-bp fragment was not detected in the patient's RNA. **b**, DGGE of PCR products corresponding to exon 19, showing the co-segregation of the mutation with the phenotype in the MRX60 family. White squares, unaffected males; black squares, affected males; white circles, unaffected females; dotted symbols, phenotypically normal carrier females.

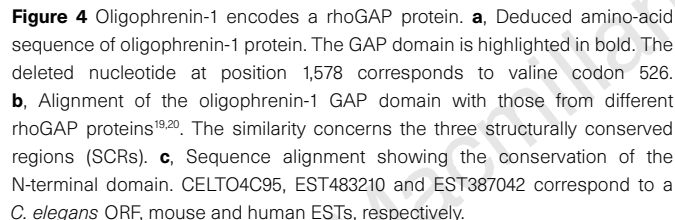
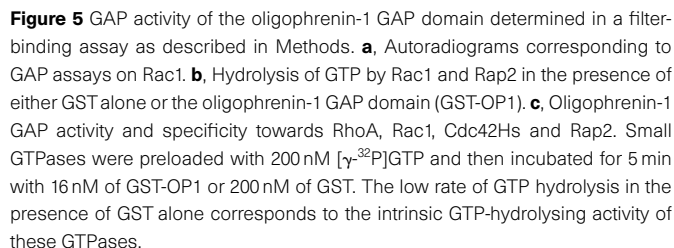


Figure 4 Oligophrenin-1 encodes a rhoGAP protein. **a**, Deduced amino-acid sequence of oligophrenin-1 protein. The GAP domain is highlighted in bold. The deleted nucleotide at position 1,578 corresponds to valine codon 526. **b**, Alignment of the oligophrenin-1 GAP domain with those from different rhoGAP proteins^{19,20}. The similarity concerns the three structurally conserved regions (SCRs). **c**, Sequence alignment showing the conservation of the N-terminal domain. CELTO4C95, EST483210 and EST387042 correspond to a *C. elegans* ORF, mouse and human ESTs, respectively.



in all parts of the developing neuroepithelium of the neural tube. During later differentiation stages and in the mature brain, a significant level of expression is visible in different brain structures (data not shown).

Our results indicate that loss of oligophrenin-1 activity in humans results in cognitive impairment. Also, the issue is raised of the functional consequences resulting from an inactivation of oligophrenin-1 towards interacting proteins and/or target GTPases, although this cannot yet be tested experimentally. Inactivation of oligophrenin-1 might affect the activity of interacting proteins or cause constitutive activation of potential rhoGTPase targets. Such constitutive activation of Rho family members affects cell migration, axon outgrowth and morphogenesis *in vivo*⁶⁻⁸. Our findings indicate that the loss of function of oligophrenin-1 may disturb, at least, on a small scale, signal transduction pathways involved in cell migration and axon outgrowth during development of the nervous system: mental retardation could be the clinical expression of such neuropathological changes. □

Methods

Case report and MRX families. Clinical data and diagnosis concerning the female patient with the t(X;12) translocation have been described⁹. Linkage studies in MRX families are summarized in Table 1. Cognitive impairment is the only common feature between patients in these families.

YAC, PAC and cosmid clones. YAC and PAC clones were obtained from the UK-HGMP Resource Centre and from the German Resource Centre. Cosmid clones were isolated from the ICRF human X-chromosome library¹⁵ using B2-cDNA as a probe. Cosmids 2C6, 4D2 and 35 are from a cosmid library corresponding to YAC 4690. Primer sequences of STSs (sequence-tagged sites) are available in the Genome Data Base. STS16T3 is a 189-bp fragment amplified by using the primers: C16T3F (5'-CACAGCAAGCAATAAGCACT-3') and C16T3R (5'-TGTCTCTGTGCTCTTTCCA-3'). Overlaps between clones and STS mapping were determined by a combination of STS/EST amplification and hybridization.

cDNA isolation. Approximately 1×10^6 recombinant clones of a λ gt10 human fetal brain cDNA library were screened using an RT-PCR product as probe, according to standard techniques¹⁶. Inserts of positive clones were subcloned into Bluescript vector and sequenced. 3' Rapid amplification of cDNA ends-PCR (RACE-PCR) was used to clone the 3' end of the cDNA.

RT-PCR and mutation analysis. For RT-PCR, the forward primer in exon 1 was: 5'-CGGCCCTGGTCAGGAAACGCGA-3'; and reverse primer in exon 2 was: 5'-TGAACGCCAGCGGGGATGACCC-3'. After electrophoresis of PCR products, Southern blots were hybridized with an internal oligonucleotide primer: 5'-GTCTCTGTGCCCCACGACCA-3'. The ubiquitously expressed transcript produced by the distal part of the dystrophin gene was used as an internal standard.

For mutation analysis, the 23 coding exons were amplified from genomic DNA with specific primers. In each amplification, one primer was a 5' psoralen-modified primer¹⁷. PCR products were checked on agarose gels before DGGE analysis and direct sequencing. Exon 19, for which DGGE analysis is shown in Fig. 3, was amplified with primers 19F (5'-GTTAATCTTGCCCCCTTTCT-3') and 19R (5' psoralen-TAGGAAGACAGG-TAGTGAGAAT), yielding a 221-bp product. Heteroduplexes formed by mixing normal and mutated PCR products were then electrophoresed through a 25–65% denaturant 6% polyacrylamide gel for 7.5 h at 60 °C and 160 V.

GAP activity. GAP activity was assayed according to standard procedures¹⁸. Briefly, GTPases proloaded with GTP were prepared by incubating 200 nM recombinant proteins with 5 μ M [γ -³²P]GTP (30 Ci mmol⁻¹) at room temperature. GTP hydrolysis, in the presence of oligophrenin 1 GAP domain or GST, was initiated by raising the final concentration of MgCl₂ and GTP and was stopped at different times with 200 μ l 'stop' buffer. Samples were blotted onto nitrocellulose using vacuum filtration and washed with stop buffer. The remaining [γ -³²P]GTP bound to GTPases was determined by scanning autoradiograms using appropriate software (Bioprofil, Vilbert Lourmat). GAP activity was assayed in the presence of decreasing amounts (280 to 3 nM) of oligophrenin-1 GAP domain. GAP activity was tested for each GTPase at least three times.

Received 31 October 1997; accepted 20 February 1998.

1. Lubs, H. A. *et al.* XLMR genes: update 1996. *Am. J. Med. Genet.* **64**, 147–157 (1996).
2. Herbst, D. S. & Miller, J. R. Nonspecific X-linked mental retardation II: the frequency in British Columbia. *Am. J. Med. Genet.* **7**, 461–469 (1980).
3. Gedeon, A. K., Donnelly, A. J., Mulley, J. C., Kerr, B. & Turner, G. How many X-linked genes for non-specific mental retardation (MRX) are there? *Am. J. Med. Genet.* **64**, 158–162 (1996).
4. Lamarche, N. & Hall, A. Gaps for rho-related GTPases. *Trends Genet.* **10**, 436–440 (1994).
5. Hildebrand, J. D., Taylor, J. M. & Parsons, J. T. An SH3 domain-containing GTPase-activating protein for Rho and Cdc42 associates with focal adhesion kinase. *Mol. Cell. Biol.* **16**, 3169–3178 (1996).
6. Zipkin, I. D., Kindt, R. M. & Kenyon, C. J. Role of a new rho family member in cell migration and axon guidance in *C. elegans*. *Cell* **90**, 883–894 (1997).
7. Luo, L., Liao, Y. J., Jan, L. Y. & Jan, Y. N. Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev.* **8**, 1787–1802 (1994).
8. Luo, L. *et al.* Differential effects of the Rac GTPase on Purkinje cell axons and dendritic trunks and spines. *Nature* **379**, 837–840 (1996).
9. Bienvenu, T. *et al.* Mapping of the X-breakpoint involved in a balanced X;12 translocation in a female with mild mental retardation. *Eur. J. Hum. Genet.* **5**, 105–109 (1997).
10. Kel, A. E. *et al.* SITEVIDEO: a computer system for functional site analysis and recognition. Investigation of the human splice sites. *Comput. Appl. BioSci.* **9**, 617–627 (1993).
11. Solovyev, V. V., Salsmov, A. A. & Lawrence, C. B. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res.* **22**, 5156–5163 (1994).
12. Lerman, L. S. & Silverstein, K. Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Meth. Enzymol.* **155**, 482–501 (1987).
13. Davies, H. R. *et al.* Androgen insensitivity with mental retardation: a contiguous gene syndrome? *J. Med. Genet.* **34**, 158–160 (1997).
14. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
15. Nizetic, D. *et al.* Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. *Proc. Natl. Acad. Sci. USA* **88**, 3233–3237 (1991).
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1982).
17. Fernandez, E. *et al.* Use of chemical clamps in denaturing gradient gel electrophoresis: application in the detection of the most frequent Mediterranean beta-thalassemic mutations. *PCR Meth. Appl.* **3**, 122–124 (1993).
18. Settleman, J., Albright, C. F., Foster, L. C. & Weinberg, R. A. Association between GTPase activators for Rho and Ras families. *Nature* **359**, 153–154 (1992).
19. Boguski, M. S. & McCormick, F. Proteins regulating Ras and its relatives. *Nature* **366**, 643–654 (1993).
20. Aelst, L. V. & D'Souza-Schorey, C. Rho GTPases and signaling networks. *Genes Dev.* **11**, 2295–2322 (1997).

Acknowledgements. We thank the members of the families for their participation in this study; F. Francis, G. Gacou and O. Dorseuil for critically reading the manuscript; and T. Brûls for cosmid clones. This work was supported in part by grants from the AFM, AP-HP, and the Fondation Jérôme Lejeune. P.B. is supported by a Ph.D. fellowship from Ministère de la Recherche.

Correspondence and requests for materials should be addressed to J.C. (e-mail: chelly@cochin.inserm.fr). The EMBL accession number for the *oligophrenin-1* sequence reported here is AJ001189.

GABA_A receptor α 4 subunit suppression prevents withdrawal properties of an endogenous steroid

Sheryl S. Smith, Qi Hua Gong, Fu-Chun Hsu, Ronald S. Markowitz, J. M. H. French-Mullen* & Xinshe Li

Department of Neurobiology and Anatomy, Allegheny University of the Health Sciences, EPPI, 3200 Henry Avenue, Philadelphia, Pennsylvania 19129, USA
* RIN Bioscience, Zeneca Pharmaceuticals, Zeneca Inc., Wilmington, Delaware 19850–5437, USA

The hormone progesterone is readily converted to 3 α -OH-5 α -pregnan-20-one (3 α ,5 α -THP) in the brains of males and females^{1,2}. In the brain, 3 α ,5 α -THP acts like a sedative^{3–5}, decreasing anxiety and reducing seizure activity, by enhancing the function of GABA (γ -aminobutyric acid)^{6–8}, the brain's major inhibitory neurotransmitter. Symptoms of premenstrual syndrome (PMS), such as anxiety⁹ and seizure^{10,11} susceptibility, are associated with sharp declines in circulating levels of progesterone and, consequently, of levels of 3 α ,5 α -THP in the brain. Abrupt discontinuation of use of sedatives such as benzodiazepines¹² and ethanol¹³ can also produce PMS-like withdrawal symptoms. Here we report a progesterone-withdrawal paradigm, designed to mimic PMS and post-partum syndrome in a rat

model. In this model, withdrawal of progesterone leads to increased seizure susceptibility and insensitivity to benzodiazepine sedatives through an effect on gene transcription. Specifically, this effect was due to reduced levels of $3\alpha,5\alpha$ -THP which enhance transcription of the gene encoding the $\alpha 4$ subunit of the GABA_A receptor. We also find that increased susceptibility to seizure after progesterone withdrawal is due to a sixfold decrease in the decay time for GABA currents and consequent decreased inhibitory function. Blockade of the $\alpha 4$ gene transcript prevents these withdrawal properties. PMS symptoms may therefore be attributable, in part, to alterations in expression of GABA_A receptor subunits as a result of progesterone withdrawal.

We first studied the possibility that cross-tolerance to progesterone/ $3\alpha,5\alpha$ -THP and other GABA-modulatory drugs might occur following withdrawal from systematically administered progesterone. Therefore we evaluated benzodiazepine (BDZ)-mediated potentiation of GABA currents using whole-cell patch-clamp techniques on acutely dissociated pyramidal neurons from CA1 hippocampus. Twenty-four hours after progesterone withdrawal, sensitivity to the normally potentiating action of the BDZ lorazepam was reduced by 70–100% ($P < 0.001$; Figs 1a, 2a). This effect

was a result of withdrawal of the GABA-modulator $3\alpha,5\alpha$ -THP, as blockade of $3\alpha,5\alpha$ -THP formation with indomethacin during progesterone exposure prevented the BDZ insensitivity observed following progesterone withdrawal (Fig. 1a). Our previous results¹⁴ indicate that this BDZ insensitivity may be due to actions at the GABA_A receptor and requires multiple withdrawal cycles, a paradigm shown to produce a persistent ethanol-withdrawal state¹³. In addition, after withdrawal of $3\alpha,5\alpha$ -THP following chronic administration of progesterone, tolerance to the GABA-modulating effects of added $3\alpha,5\alpha$ -THP was observed (Fig. 1b; $P < 0.05$), demonstrating an additional similarity between this neuroactive steroid and the BDZs. Cross-tolerance of ethanol and $3\alpha,5\alpha$ -THP with BDZs occurs *in vitro*^{15–17}, and $3\alpha,5\alpha$ -THP has also been shown to substitute for BDZs in a drug-discrimination task¹⁸. Finally, BDZ insensitivity has been reported in women with PMS¹⁹, indicating possible clinical relevance for this suggested similarity between these GABA-modulating compounds.

Evaluation of the kinetics of GABA currents (Figs 1c, 2a) revealed a sixfold decrease in the decay time constant (τ) for GABA currents ($P < 0.01$) after withdrawal of progesterone ($\tau = 916.9 \pm 70.09$ ms) compared with control values ($\tau = 5,794.4 \pm 1515.7$ ms; Fig. 2a).

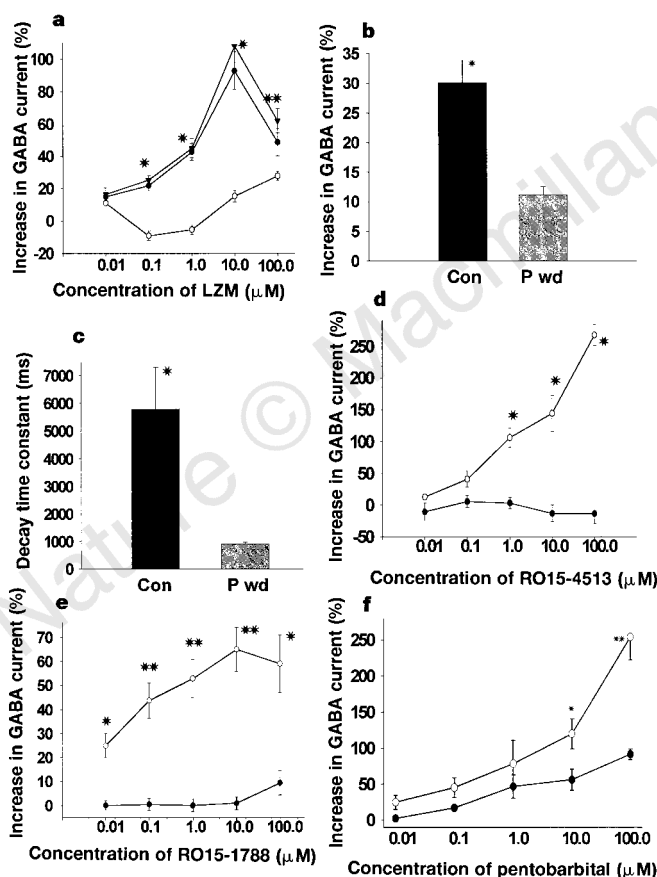


Figure 1 Profile of altered GABA_A pharmacology following 'withdrawal' of progesterone. We show concentration–response curves from averaged data demonstrating changes in GABA_A currents after withdrawal from progesterone (P) compared with controls (open circles, P withdrawal; filled circles, sham control; triangles, withdrawal from P plus 0.1 mg kg⁻¹ indomethacin, injected i.p.). **a**, Lorazepam (LZM)-mediated modulation of GABA_A currents. **b**, $3\alpha,5\alpha$ -THP (10 nM) mediated modulation of GABA_A current. Con, sham control; P wd, progesterone withdrawal. **c**, Decay time constant for GABA_A currents. **d**, Effects of the partial inverse agonist RO15-4513 on GABA_A currents. **e**, Modulation of GABA_A currents by the BDZ antagonist RO15-1788. **f**, Pentobarbital-mediated modulation of GABA_A currents; 16–26 cells tested per group, 3–6 cells tested per animal. Single asterisk indicates $P < 0.001$; double asterisks indicate $P < 0.05$.

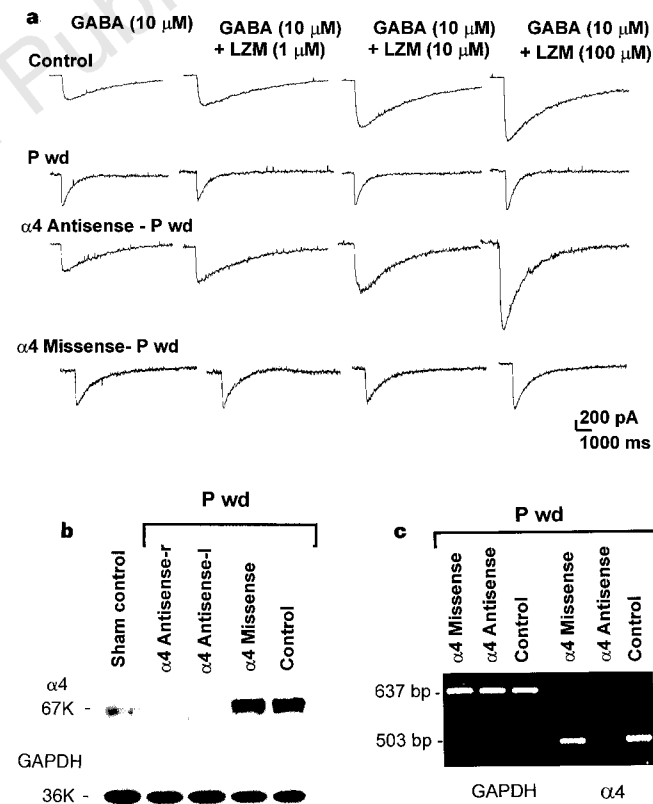


Figure 2 $\alpha 4$ antisense oligonucleotide prevents the withdrawal properties of progesterone. **a**, GABA-gated currents from the CA1 hippocampus of control animals or after progesterone withdrawal (P wd) in non-cannulated animals (second row) and after $\alpha 4$ antisense (third row) or $\alpha 4$ missense (bottom row) treatment. These results are representative of those from 20–25 cells from 12–18 rats, each. LZM, lorazepam. **b**, **c**, Oligonucleotide effects on $\alpha 4$ subunit protein and mRNA levels from animals used in **a**. -r, Right hippocampus; -l, left hippocampus. Controls include sham-implanted untreated animals (sham control) and progesterone-withdrawn non-cannulated animals (P wd control).

Such a marked decrease in GABA current decay time would markedly reduce total GABA synaptic current, which in turn would increase neuronal excitability. Behaviourally, this might be expected to result in increased anxiety, as previously found following progesterone withdrawal in rats²⁰ and during PMS⁹.

As increases in seizure activity can result from decreased GABA currents, we tested chemically induced seizure intensity after progesterone withdrawal (Fig. 3a). Eightfold higher seizure scores (200 ± 20) were found compared with controls (25 ± 5). Increased seizure activity is seen during 'withdrawal' from endogenous progesterone in susceptible women (that is, catamenial epilepsy)^{10,11} and is also a withdrawal property of other GABA-modulatory drugs such as ethanol¹³.

To determine a possible mechanism for these withdrawal effects, we characterized BDZ- and barbiturate-mediated modulation of GABA-gated currents 24 h after progesterone withdrawal. As shown in Fig. 1a, d–f, the pharmacological profile is identical to that for oocytes transfected with GABA_A receptors containing the $\alpha 4$ subunit ($\alpha 4\beta 1\gamma 2$ receptors)²¹. BDZ-insensitive GABA receptors containing the $\alpha 4$ subunit²² exhibit increased responses to the partial inverse agonist RO15-4513 (ref. 21). Here, RO15-4513 produced potent agonist-like actions (Fig. 1d), increasing GABA-gated currents up to $268 \pm 16.8\%$ after progesterone withdrawal (compared with $5.79 \pm 9.81\%$ in controls, $P < 0.001$, 10^{-4} M; Fig. 1d). In addition, progesterone withdrawal resulted in significant attenuation ($P < 0.05$) of the ability of a BDZ inverse agonist (β -CC) to inhibit GABA currents (data not shown) and converted the BDZ antagonist RO15-1788 to a potent agonist (Fig. 1e). Barbiturate-mediated potentiation of GABA currents was enhanced by two- to threefold following progesterone withdrawal (Fig. 1f).

As the effects observed by 24 h after progesterone withdrawal mimic properties reported for $\alpha 4$ -containing GABA_A receptors in oocytes²¹, we tested the possibility that progesterone withdrawal was associated with increases in hippocampal levels of the $\alpha 4$ subunit of the GABA_A receptor, which, in the CA1 hippocampus, is located mainly at the pyramidal cell²². Levels of the $\alpha 4$ subunit were significantly higher ($P < 0.05$) 8 h after progesterone withdrawal than in controls and reached maximum values by 24 h after progesterone withdrawal (Fig. 4a, c). In contrast, no change in levels of the $\gamma 2$ subunit protein were seen across the withdrawal period (Fig. 4a, c).

We used semiquantitative reverse transcription with polymerase chain reaction (RT-PCR) techniques to assess levels of $\alpha 4$ messenger RNA. Levels of $\alpha 4$ mRNA were increased (Fig. 4b, c) by 24 h after progesterone withdrawal compared with controls. Similar increases in $\alpha 4$ mRNA or protein levels have been reported after withdrawal or chronic exposure to ethanol^{23,24} or BDZs²⁵, respectively.

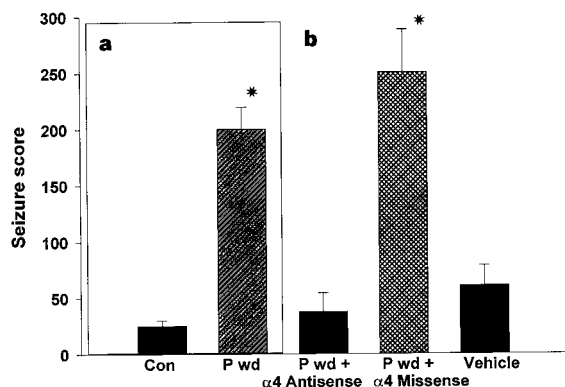


Figure 3 Effects of progesterone withdrawal on seizures. **a**, Increased seizure-like activity is observed following progesterone withdrawal (P wd) compared with controls (Con). **b**, This effect was blocked by $\alpha 4$ antisense oligonucleotide; $n = 6$ animals per group, asterisk indicates $P < 0.05$ compared with all other groups.

We prevented the observed increases in $\alpha 4$ subunit levels with $\alpha 4$ antisense oligonucleotide administration *in vivo* (Fig. 2b, c). This treatment removed the BDZ insensitivity normally found at 24 h after progesterone withdrawal (Fig. 2a). (Dose range of $\alpha 4$ antisense tested: 0, 0.1, 0.5, 1.0, 2.5, 5.0 μ g per day; 1.0 μ g per day was a maximal dose and 0.5 μ g per day produced a response that was $\sim 50\%$ of maximal, for both $\alpha 4$ subunit suppression and prevention of the withdrawal properties of progesterone; $n = 6$ animals per group, $P < 0.05$ versus 0 μ g per day.) In contrast, treatment with $\alpha 4$ missense (Fig. 2a) or infusion of $\alpha 1$ antisense ($n = 18$; data not shown) did not alter this parameter. These results indicate that the increase in $\alpha 4$ subunit observed 24 h after progesterone withdrawal may be a necessary and sufficient contributing factor in producing cross-tolerance to BDZ potentiation of GABA currents.

The decrease in GABA current decay time following progesterone withdrawal was also prevented by $\alpha 4$ antisense treatment (Fig. 2a), indicating that increases in $\alpha 4$ subunit levels also alter intrinsic GABA/Cl⁻ channel properties, resulting in decreased total GABA currents.

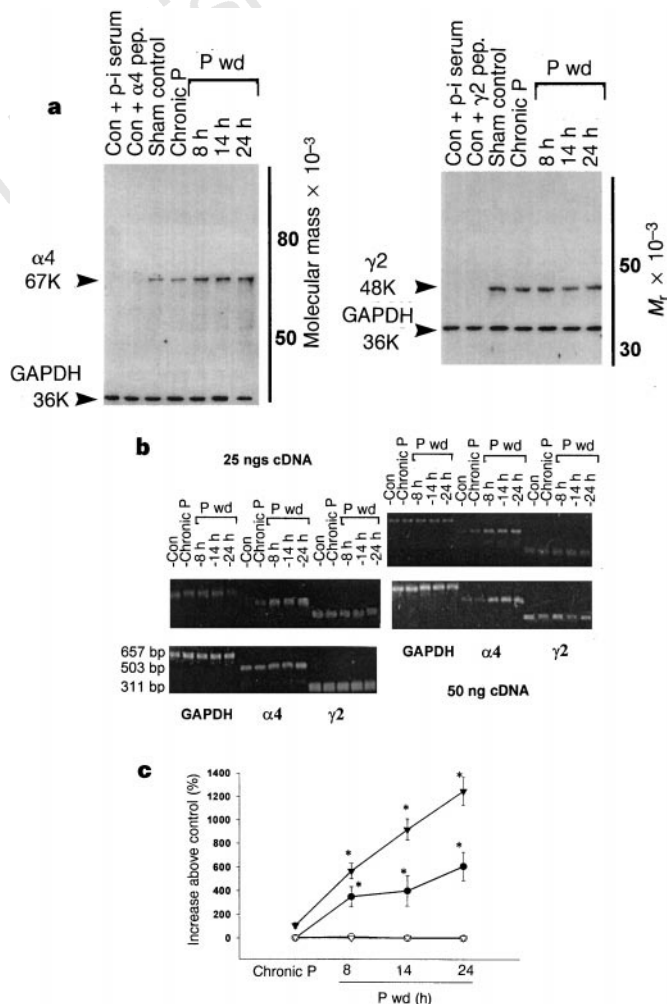


Figure 4 Progesterone withdrawal increases protein and mRNA levels of the GABA_A receptor $\alpha 4$ subunit. P, progesterone. Con, control; pep., peptide; p-i, pre-immune. **a**, Western blot analysis of $\alpha 4$ (67K) (left) and $\gamma 2$ (48K) (right) subunits, with a GAPDH (36K) control protein. **b**, Semiquantitative RT-PCR analysis of $\alpha 4$, $\gamma 2$ and GAPDH mRNA levels. Results are representative of those obtained after 2–3 replications at 25–27 cycles (25 ng cDNA; left) and 23–25 cycles (50 ng cDNA; right). Staining of bands is in the linear range across cycle and concentration ranges. **c**, Summary data from two to three replications for six animals per group. Filled circles, $\alpha 4$ mRNA; open circles, $\gamma 2$ mRNA; filled triangles, $\alpha 4$ protein; open triangles, $\gamma 2$ protein. Asterisk indicates $P < 0.05$ compared with control.

The proconvulsant withdrawal property of progesterone, most likely a result of this decreased inhibitory current, is also attributable to increased levels of the $\alpha 4$ GABA receptor subunit. Pretreatment with $\alpha 4$ antisense, but not missense, oligonucleotides prevented the increased seizure susceptibility observed after progesterone withdrawal (Fig. 3b) where seizure score was not significantly different from sham controls.

Our results indicate that fluctuations in endogenous progesterone levels may result in plasticity of the GABA_A receptor through the GABA-modulating $3\alpha,5\alpha$ -THP. Fluctuations in levels of these neuroactive steroids are associated with the menstrual and pregnancy cycles, and are induced by stress in males². Manipulation of GABA_A receptor $\alpha 4$ subunit levels may prevent cross-tolerance with sedative drugs and reduce generalized excitability and increased seizure susceptibility^{10,11} associated with periods of endogenous progesterone withdrawal. □

Methods

Progesterone withdrawal. Progesterone was administered in capsules using a paradigm shown to result in physiological levels of both progesterone¹ and $3\alpha,5\alpha$ -THP (ref. 14). Capsules were replaced weekly, with a two-day interruption, for three weeks (multiple withdrawal schedule)¹⁴ and tested 8–24 h later on proestrus⁷. Controls received empty capsules.

Electrophysiology. GABA-gated Cl^- currents were analysed using a whole-cell patch-clamp procedure which included an ATP-regeneration system¹⁴ and a rapid perfusion system for drug delivery¹⁴. (Drugs from RBI biochemicals.) Statistical methods included analysis of variance (ANOVA) with Student-Newman-Keuls post-hoc procedures. Current kinetics were analysed with Origin Software (Microcal Inc.); the inactivation time course was approximated by a single first-order exponential function²⁶.

Western blot analysis. Hippocampal membranes electrophoresed onto 9% SDS–polyacrylamide gels were transferred to a polyvinylidene difluoride membrane (Bio-Rad). The $\alpha 4$ subunit was detected with a rabbit antibody against a peptide of rat $\alpha 4$ (ref. 27) (amino acids 517–523, with an amino-terminal cysteine) as a band of relative molecular mass (M_r) 67K (Genosys, Texas). $\gamma 2$ subunits were detected with antibodies directed against amino acids 316–352 of the peptide as a 48K band²⁸. Negative controls included the use of pre-immune serum and blockade with $50 \mu\text{g ml}^{-1}$ $\alpha 4$ or $\gamma 2$ peptide (Genosys Inc.). Membranes were incubated for 1–2 h at room temperature with a 1:10,000 ($\alpha 4$) or 1:3,000 ($\gamma 2$) and 1:20,000 (glyceraldehyde-3-phosphate dehydrogenase, GAPDH) dilution of the antibody, followed by a 1:5,000 dilution of horseradish-peroxidase-conjugated donkey anti-rabbit IgG (Amersham). $\alpha 4$ and $\gamma 2$ immunoreactive band densities from the gel were quantified using a CCD camera (Umax Scanner) and densitometry with One-Dscan software after standardization with a control antibody (GAPDH, Chemicon).

Semi-quantitative RT-PCR. We used oligonucleotide primers²⁹ for the $\alpha 4$ (503 base pairs (bp)) and $\gamma 2$ (311 bp) subunits of the GABA_A receptor (Operon Inc.) and a control primer, for GAPDH (657 bp). Total RNA was isolated³⁰ and subjected to reverse transcription (SuperScript II Rnase RT, GibcoBRL). Complementary DNA containing identical amounts of total RNA was subjected to PCR amplification for 23–30 cycles³⁰ using two different concentrations of cDNA (25 or 50 ng). Following gel electrophoresis (FMC Bioproducts), the density of DNA bands stained with ethidium bromide (single band per PCR product) was quantified as described, standardized to the GAPDH control³⁰. Values were averaged from both cDNA concentrations at the lower cycle number known to be in the linear range.

Antisense administration. $\alpha 4$ and $\alpha 1$ antisense or missense oligodeoxynucleotides, phosphorothioated at all positions and purified by HPLC (Genosys Inc.) were as follows: $\alpha 4$ antisense, 5'-ACCTTCTGGACAGAAACC-3' (+3, relative to the ATG for initiating translation); $\alpha 4$ missense, 5'-ACCATCTA-GACGGAATCC-3'; $\alpha 1$ antisense, 5'-AGACCCCGACTTTTCTTC-3'. (Missense oligonucleotides were constructed with every fourth base scrambled, yielding an identical GC content to the $\alpha 4$ antisense.) Oligonucleotides were delivered by a subcutaneously implanted osmotic minipump (2001, Alza Corp.) at a concentration of 0.1–5.0 $\mu\text{g per day}$ (vehicle, 0.35% BSA/0.15 M saline), 1 $\mu\text{l per h}$, for 48–72 h through 29-gauge tubing attached to a previously implanted cannula which terminated within the right lateral

ventricle. Successful reduction in $\alpha 4$ and $\alpha 1$ subunit protein/mRNA levels was verified using western blot and semiquantitative RT-PCR techniques.

Electrophysiological assessments were performed on pyramidal neurons freshly isolated from the CA1 region of right hippocampus. Controls received vehicle, $\alpha 4$ missense or $\alpha 1$ antisense oligonucleotides (1 $\mu\text{g per day}$) delivered through a right lateral ventricle cannula.

Seizure induction. Seizure activity was precipitated with intraperitoneal (i.p.) injection of the inverse agonist n-methyl- β -carboline-3-carboxamide (β -CC, 26 mg kg^{-1} , i.p. in 45% w/v 2-hydroxypropyl- β -cyclodextrin/saline). Rats were scored on latency to seizure and seizure severity and duration, and a seizure score was calculated with the formula $\Sigma(\text{seizure severity rating}) (\text{duration, min})/\text{latency to seizure (min)}$ using a scale from 1 to 4. Differences between groups were calculated using ANOVA plus Student-Newman-Keuls post-hoc procedures.

Received 8 December 1997; accepted 13 February 1998.

- Smith, S. S. Female sex steroids: from receptors to networks to performance—actions on the sensorimotor system. *Prog. Neurobiol.* **44**, 55–86 (1994).
- Purdy, R. H., Morrow, A. L., Moore, P. H. Jr & Paul, S. M. Stress-induced elevations of GABA type A receptor-active steroids in the rat brain. *Proc. Natl Acad. Sci. USA* **88**, 4553–4557 (1991).
- Bitran, D., Hilvers, R. J. & Kellogg, C. K. Anxiolytic effects of 3α -OH- 5α [β]-pregnan-20-one: endogenous metabolites of progesterone that are active at the GABA_A receptor. *Brain Res.* **561**, 157–161 (1991).
- Brot, M. D., Akwa, Y., Purdy, R. H., Koob, G. F. & Britton, K. T. The anxiolytic-like effects of the neurosteroid allopregnanolone—interactions with GABA_A receptors. *Eur. J. Pharmacol.* **325**, 1–7 (1997).
- Belelli, D., Bolger, M. B. & Gee, K. Anticonvulsant profile of the progesterone metabolite 5α -pregnan- 3α -ol-20-one. *Eur. J. Pharmacol.* **166**, 325–329 (1989).
- Majewska, M. D., Harrison, N. L., Schwartz, R. D., Barker, J. L. & Paul, S. M. Steroid hormone metabolites are barbiturate-like modulators of the GABA receptor. *Science* **232**, 1004–1007 (1986).
- Smith, S. S., Waterhouse, B. D., Chapin, J. K. & Woodward, D. J. Progesterone alters GABA and glutamate responsiveness: a possible mechanism for its anxiolytic action. *Brain Res.* **400**, 353–359 (1987).
- Twyman, R. E. & MacDonald, R. L. Neurosteroid regulation of GABA_A receptor—single channel kinetic properties of mouse spinal cord neurons in culture. *J. Physiol. (Lond.)* **456**, 215–245 (1992).
- Dennerstein, L. et al. Progesterone and premenstrual syndrome: a double blind crossover trial. *Brit. Med. J.* **290**, 1617–1621 (1985).
- Backstrom, T., Zetterlund, B., Blom, S. & Romano, M. Effects of intravenous progesterone infusions on the epileptic discharge frequency in women with partial epilepsy. *Acta Neurol. Scand.* **69**, 240–248 (1984).
- Herzog, A. G. Progesterone therapy in women with complex partial and secondary generalized seizures. *Neurology* **45**, 1660–1662 (1995).
- File, S. E. The history of benzodiazepine dependence: a review of animal studies. *Neurosci. Biobehav. Rev.* **14**, 135–146 (1990).
- Kokka, N., Sapp, D. W., Taylor, A. M. & Olsen, R. W. The kindling model of alcohol dependence: similar persistent reduction in seizure threshold to pentylenetetrazol in animals receiving chronic ethanol or chronic pentylenetetrazol. *Alcohol Clin. Exp. Res.* **17**, 525–531 (1993).
- Costa, A.-M. N., Spence, K. T., Smith, S. S. & Ffrench-Mullen, J. M. H. Withdrawal from the endogenous steroid progesterone results in GABA_A currents insensitive to benzodiazepine modulation in rat CA1 hippocampus. *J. Neurophys.* **74**, 464–469 (1995).
- Friedman, L., Gibbs, T. T. & Farb, D. H. γ -Aminobutyric acid_A receptor regulation: Chronic treatment with pregnanolone uncouples allosteric interactions between steroid and benzodiazepine recognition sites. *Mol. Pharmacol.* **44**, 191–197 (1993).
- Buck, K. J. & Harris, R. A. Benzodiazepine agonist and inverse agonist actions on GABA_A receptor-operated chloride channels. II. Chronic effects of ethanol. *J. Pharmacol. Exp. Ther.* **253**, 713–719 (1990).
- Yu, X. J. & Ticku, M. K. Chronic neurosteroid treatment produces functional heterologous uncoupling at the gamma-aminobutyric acid type A/benzodiazepine receptor complex in mammalian cortical neurons. *Mol. Pharmacol.* **47**, 603–610 (1994).
- Ator, N. A., Grant, K. A., Purdy, R. H., Paul, S. M. & Griffiths, R. R. Drug discrimination analysis of endogenous neuroactive steroids in rats. *Eur. J. Pharmacol.* **241**, 237–243 (1993).
- Sundstrom, L., Ashbrook, D. & Backstrom, T. Reduced benzodiazepine sensitivity in patients with premenstrual syndrome—a pilot study. *Psychoneuroendo* **22**, 25–38 (1997).
- Gallo, M. A. & Smith, S. S. Progesterone withdrawal decreases latency to and increases duration of electrified prod burial: a possible rat model of PMS anxiety. *Pharmacol. Biochem. Behav.* **46**, 897–904 (1993).
- Wafford, K. A. et al. Functional characterization of human GABA receptors containing the $\alpha 4$ subunit. *Mol. Pharmacol.* **50**, 670–678 (1996).
- Wisden, W. et al. Cloning, pharmacological characteristics and expression pattern of the rat GABA_A receptor $\alpha 4$ subunit. *FEBS Lett.* **289**, 227–230 (1991).
- Mahmoudi, M., Kang, M. H., Tillakaratne, N., Tobin, A. J. & Olsen, R. W. Chronic intermittent ethanol treatment in rats increases GABA_A receptor $\alpha 4$ subunit expression—possible relevance to alcohol dependence. *J. Neurochem.* **68**, 2485–2492 (1997).
- Devaud, L. L., Fritschy, J. M., Sieghart, W. & Morrow, A. L. Bidirectional alterations of GABA_A receptor subunit peptide levels in rat cortex during chronic ethanol consumption and withdrawal. *J. Neurochem.* **69**, 126–130 (1997).
- Holt, R. A., Bateson, A. N. & Martin, I. L. Chronic treatment with diazepam or abecarnil differently affects the expression of GABA_A receptor subunit mRNAs in the rat cortex. *Neuropharmacol.* **35**, 1457–1463 (1996).
- Lavoie, A. M., Tingey, J. J., Harrison, N. L., Pritchett, D. B. & Twyman, R. E. Activation and deactivation rates of recombinant GABA(A) receptor channels are dependent on alpha-subunit isoform. *Biophys. J.* **73**, 1–9 (1997).
- Kern, W. & Sieghart, W. Polyclonal antibodies directed against an epitope specific for the $\alpha 4$ -subunit of GABA_A receptors identify a 67-kDa protein in rat brain membranes. *J. Neurochem.* **62**, 764–769 (1994).
- Mossier, B., Togel, M., Fuchs, K. & Sieghart, W. Immunoaffinity purification of GABA_A receptors containing $\gamma 2$ subunits. *J. Biol. Chem.* **269**, 25777–25782 (1994).

29. Tyndale, R. F., Hales, T. G., Olsen, R. W. & Tobin, A. J. Distinctive patterns of GABA_A receptor subunit mRNAs in 13 cell lines. *J. Neurosci.* **14**, 5417–5428 (1994).
30. Lau, E. C., Li, Z.-Q., Santos, V. & Slavkin, H. C. Messenger RNA phenotyping for semi-quantitative comparison of glucocorticoid receptor transcript levels in the developing embryonic mouse palate. *J. Steroid Biochem. Mol. Biol.* **46**, 751–758 (1993).

Acknowledgements. We thank I. Fischer for assistance; Hoffman-La Roche for the RO15-1788 sample; D. S. Faber, M. M. McCarthy and D. H. Smith for critical reading of the manuscript; and D. O'Reilly for editorial comments. The work was supported by a grant from the NIH.

Correspondence and requests for materials should be addressed to S.S.S. (e-mail: smiths@auhs.edu).

Modifying the mechanical property and shear threshold of L-selectin adhesion independently of equilibrium properties

Kamal D. Puri[†]*, Shuqi Chen^{*} & Timothy A. Springer^{*}

^{*} The Center for Blood Research and Harvard Medical School, Department of Pathology, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

Interactions between adhesion molecules on two different cells differ from interactions between receptors and soluble ligands in that the adhesion molecule interaction (bond) is often subjected to force. It is widely assumed by cell biologists that the 'strength' of a bond is a simple function of the affinity of one adhesion molecule for the other, whereas biophysicists suggest that bonds have 'mechanical properties' that affect their strength. Mechanical properties are a function of the shape of the energy landscape related to bond formation and dissociation, whereas affinity is related only to the net energy change^{1–5}. Mechanical properties determine the amount by which the kinetics and affinity of bonds are altered by applied force. To date there has been no experimental manipulation of an adhesion molecule that has been shown to affect mechanical properties. L-selectin is an adhesion molecule that mediates lymphocyte binding to, and rolling on, high endothelial venules; these are prerequisites for the emigration of lymphocytes from the bloodstream into lymph nodes. Here we report a selective and reversible chemical modification of a mucin-like ligand that alters the mechanical properties of its bond with L-selectin. The effect of force on the rate of bond dissociation, that is, on a mechanical property, is altered, whereas there is little or no effect of the modification on the rate of bond dissociation in the absence of force. Moreover, the puzzling requirement for hydrodynamic shear flow above a threshold level for L-selectin interactions^{6–9} is dramatically altered.

L-selectin is expressed on leukocytes and binds to a sialyl-Lewis-X-like carbohydrate ligand that is expressed on mucin-like molecules on the endothelium and on other leukocytes^{10,11}. L-selectin is important in initiating rolling interactions between leukocytes in the blood and endothelial cells of the blood-vessel wall. The presence of a small number of L-selectin–ligand interactions at any one time, and rapid bond breakage and formation, allow the zone of adhesive contact to be translated along the vessel wall as cells roll in response to hydrodynamic drag forces^{12–14}. Like all other adhesive interactions, those through selectins can be inhibited from forming, or reversed, by applied force. As shear stress on the vessel wall, a stress which is proportional to the force on a cell, is increased above a certain level, the binding of cells in the bloodstream to the vessel wall can no longer be initiated (Fig. 1a). Cells bound to the vessel wall at moderate wall shear stresses are dislodged at high wall

shear stresses (Fig. 1b). It is puzzling that adhesive interactions through L-selectin do not occur at low shear stresses^{6,8,13}. Rolling adhesions through L-selectin are not initiated at stress levels below about 0.4 dyn per cm² (Fig. 1a), and cells rolling at moderate wall shear stresses are detached when shear is decreased to <0.4 dyn per cm² (Fig. 1b).

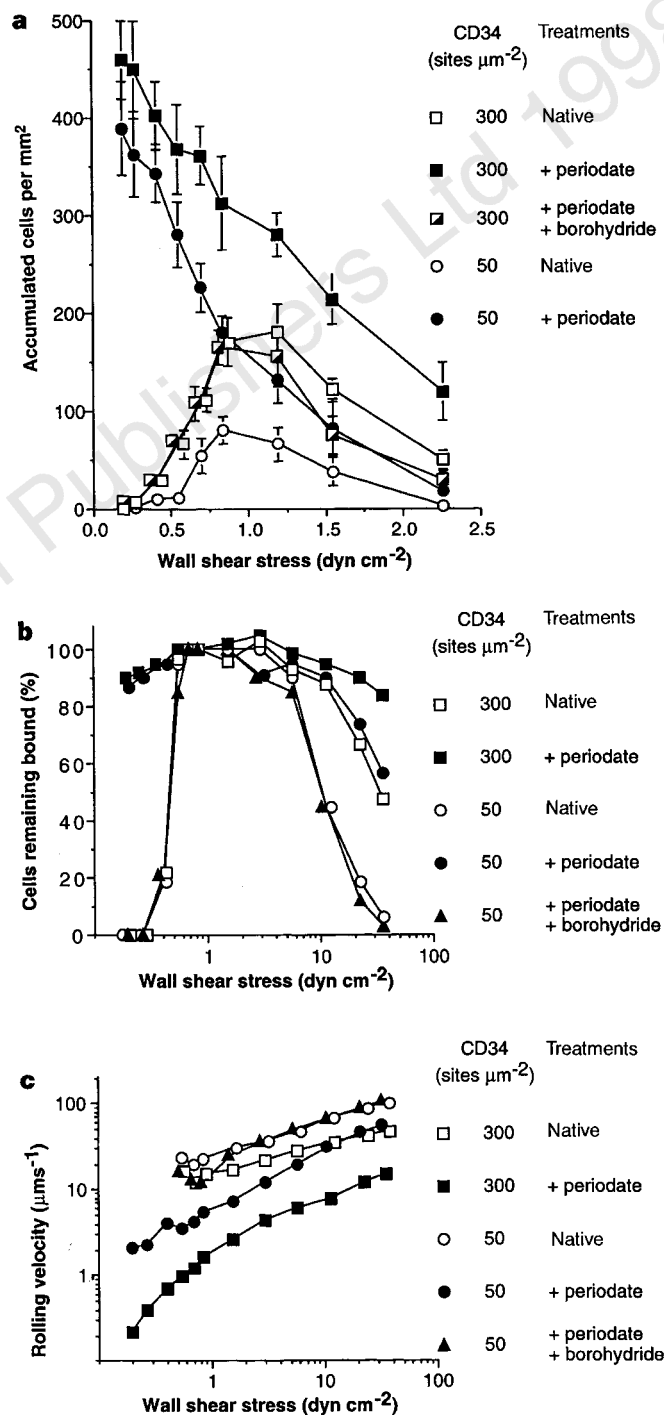


Figure 1 Mild periodate treatment of the L-selectin counter-receptor CD34 abolishes the requirement for a threshold of shear stress for L-selectin-mediated adhesion. This abolition is independent of the strength of rolling adhesions. Substrates bearing the CD34 component of peripheral-node addressin were or were not treated with mild periodate, and were or were not later treated with borohydride. **a**, Accumulation of lymphocytes at different wall shear stresses. **b**, Resistance of lymphocytes accumulated on the substrate at 0.84 dyn per cm² to detachment by subsequent increases or decreases in wall shear stress. **c**, Rolling velocity.

[†] Present address: CytoTherapeutics Inc., Protein Chemistry, 701 George Washington Highway, Lincoln, Rhode Island 02865, USA.

Chemical modification of enzymes has previously been used to demonstrate that K_M and k_{cat} are separable properties. Mild treatment of L-selectin ligands with periodate increases their binding to L-selectin^{15,16} and we tested whether mechanical properties of the interaction were affected. This chemical modification is highly selective for sialic acid; the exocyclic C8 and C9 carbons are cleaved and the C7 hydroxyl is converted to an aldehyde. We purified CD34, one of the mucin-like ligands for L-selectin^{17,18}, from tonsil. CD34 was then adsorbed to the lower wall of a flow chamber. Leukocyte interactions with the same CD34 substrate were measured before and after treatment of the substrate with mild periodate, and in some cases again after reduction of the sialic acid C7 aldehyde to a hydroxyl by using borohydride. We did not expose leukocytes to any of the reagents and removed them from the flow chamber during chemical treatments of the CD34 substrate. Periodate increased the strength of rolling interactions, as measured by resistance to detachment and slowed rolling velocity, much as did a six- to eightfold increase in CD34 density (Fig. 1b, c)¹⁶. These effects were reversed by reduction of the C7 aldehyde by borohydride. Interactions were specific, because they were completely inhibited by a monoclonal antibody against L-selectin, by EDTA, and by fucoidan (data not shown)¹⁶. The requirement for shear stress at levels above ~ 0.4 dyn per cm^2 for adhesion through L-selectin⁶ was abolished by periodate treatment (Fig. 1a, b). This abolition of the shear threshold was seen even when periodate-treated substrates at 50 sites per μm^2 and native substrates at 300 sites per μm^2 were compared (Fig. 1a); these substrates were similar in terms of the strength of adhesive interactions (Fig. 1b), the rolling velocity (Fig. 1c), and the number of rolling cells that accumulated at 0.8 dyn per cm^2 and above (Fig. 1a). Cells that were accumulated at 0.8 dyn per cm^2 and were rolling on native substrates or substrates treated with periodate and borohydride detached when shear stress was reduced to <0.4 dyn per cm^2 (Fig. 1b). In contrast, cells on

periodate-treated substrates remained rollingly adherent when shear was reduced (Fig. 1b). Therefore, periodate treatment dramatically and reversibly alters adhesive interactions with L-selectin at levels of stress <0.8 dyn per cm^2 , independently of the overall strength of adhesive interactions above 0.8 dyn per cm^2 .

At densities of selectins or their ligands on vessel walls that are too low to support rolling, cells are transiently tethered to vessel walls, that is, they associate and then dissociate^{13,14}. The kinetics of cellular dissociation are first-order and have other properties consistent with dissociation of single bonds^{13,14}, including agreement with kinetics measured for purified molecules¹⁹. The mechanical properties of the tether bond are measurable as the amount by which force on the cell, and hence on the tether bond, increases the dissociation rate constant, k_{off} ^{13,14}. The kinetics of transient tethers to CD34 were measured using automated image analysis. The first 90–99% of the transient tethers to dissociate fit a straight line, showing first-order dissociation kinetics (Fig. 2a, b). The remaining 1–10% of cells dissociated more slowly (not shown), either because of background problems or because of formation of more than one tether. The transient-tether-dissociation rate constant (k_{off}) was independent of CD34 density over the range of 0.5–10 sites per μm^2 on both native and periodate-treated substrates (not shown)¹³.

To quantify the mechanical properties of the receptor–ligand bond, we plotted k_{off} as a function of force on the bond (F_b) (Fig. 3). The relationship between wall shear stress and F_b is linear and has previously been estimated^{13,14}. The increase in k_{off} as a function of force was greater for the native ligand than for the periodate-treated

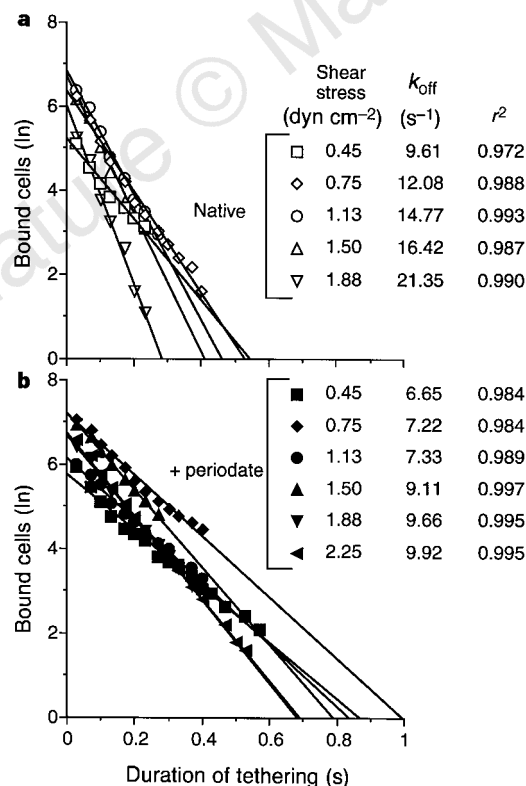


Figure 2 The kinetics of dissociation of transiently tethered neutrophils from native and mild-periodate-treated CD34. Kinetics were measured at different wall shear stresses at 0.5 sites per μm^2 of native CD34 (**a**) and mild-periodate-treated CD34 (**b**). r , coefficient of correlation.

	$k_{off}(\text{s}^{-1})$	$\sigma(\text{\AA})$	χ^2	d.f.
Native	8.0 ± 0.3	0.18 ± 0.01	3.47	32
+ periodate	5.6 ± 0.2	0.09 ± 0.01	3.56	38
+ periodate + borohydride	5.9 ± 0.6	0.25 ± 0.02	2.79	11

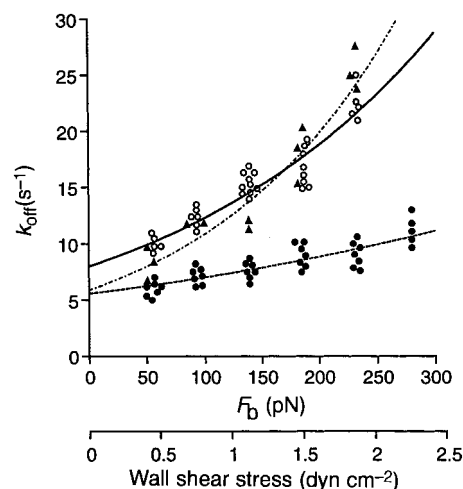


Figure 3 The effect of mild periodate and borohydride treatment on the dissociation kinetics of transient tethers between L-selectin and its ligand and on the mechanical strength of L-selectin–ligand interactions. k_{off} was determined at CD34-site densities from 0.5 to 10 sites per μm^2 . Force on the bond, F_b , was estimated using the length of the lever arm between the tether point on the substrate and the projected centre of the cell on the substrate, as described¹³. Each group of points represents a single wall shear stress; points were separated horizontally so all can be seen. Lines show the fit to Bell's model¹. The values of k_{off} , the bond separation distance (σ), the total χ^2 value, and degrees of freedom (d.f.) are shown. A fit to a Hookean spring model² yielded κ/F_x values (the bond spring constant divided by the fraction of bond stress devoted to dissociation) of 8.6 ± 0.7 , 17.9 ± 2.0 , and $5.9 \pm 0.6 \text{ N m}^{-1}$, for native, periodate-, and periodate- and borohydride-modified substrates, respectively, and similar χ^2 values.

ligand (Fig. 3). The mechanical effect of mild periodate was almost completely reversed by reduction with borohydride (Fig. 3). The data were fitted to two theoretical models relating k_{off} to F_b (refs 1, 2). As measured with the exponential constants σ in the Bell model¹ (Fig. 3) or κ/f_k in the spring model² (Fig. 3 legend), the mechanical strength of the tether bond is doubled when an aldehyde is present at the 7-position of sialic acid, compared with either the native structure or the mildly oxidized and reduced structure, which both have a hydroxyl at the 7-position.

The overall mechanical strength of a bond will be a function of the effect of force on both k_{on} and k_{off} . The frequency of transient tethers is related to k_{on} , although not necessarily linearly²⁰. Periodate also seemed to increase the mechanical strength of the tether bond as measured by transient-tether frequency, because, as shear stress was increased from 0.75 to 1.88 dyn per cm^2 , the frequency of tethers to the native substrate fell by 6.5-fold but the frequency of tethers to the periodate-treated substrate fell by only 2.6-fold (Fig. 4).

We have shown that the mechanical properties of a receptor–ligand bond can be altered by chemical modification of the ligand, independently of effects on k_{off} in the absence of force. Thus, the chemical properties that govern mechanical strength and affinity are distinct, analogous to those that govern k_{cat} and K_M for enzymes. Adhesion molecules may therefore differ from one another in both mechanical strength and affinity, and, to understand the function of adhesion molecules in resisting forces *in vivo*, both properties must be measured. The details of how chemistry regulates the mechanical properties of L-selectin ligands are currently unknown. An aldehyde oxygen is more basic than a hydroxyl oxygen, and therefore is a better acceptor for hydrogen bonds²¹. Sialic acid is a key part of the sialyl Lewis-X structure recognized by L-selectin, but it is also present on non-ligand-bearing glycans attached throughout the length of the heavily glycosylated mucin-like region of D34. Mild periodate treatment has less effect on rolling interactions involving sialyl Lewis-X glycolipid than on interactions involving CD34 (ref. 16).

The mechanical properties of a bond can be altered by modifying the elasticity of regions outside the receptor–ligand interface²²; therefore, the C7 aldehyde group on sialic acid might be able to form hydrogen bonds that alter the elastic spring constant of the

mucin-like region of CD34. This might explain the effect on mechanical strength and not on k_{off} in the absence of force. The reversible alteration of the shear-threshold requirement by mild periodate is remarkable, particularly as the same threshold exists over a range of CD34 densities, and as substrates that give equivalent rolling velocities and resistance to detachment at stress levels above 0.8 dyn per cm^2 nonetheless behave differently from each other at <0.8 dyn per cm^2 . We suggest that the elasticity of the mucin-like region of CD34 may be important in its mechanical properties, and it is interesting that ligands that lack mucin-like regions, fucoidan and sialyl Lewis-X glycolipid do not exhibit a shear threshold⁶. Although the explanation for the shear threshold remains unclear, our data indicate that there may be an important connection between the shear-threshold phenomenon and the mechanical properties of the L-selectin–ligand bond. Furthermore, our results emphasize the importance of mechanical properties for regulating cell adhesion in shear flow. □

Methods

Flow assays. The CD34 component of human peripheral-node addressin was isolated from detergent lysates of human tonsils by successive immunoaffinity chromatography on MECA-79 monoclonal antibody/Sepharose and CD34 monoclonal antibody/Sepharose¹⁸. Drops of 50 μl CD34 (350 to 60 ng ml^{-1} for rolling experiments and 12 to 0.6 ng ml^{-1} for transient-tethering experiments, in 0.01 M Tris HCl, 0.15 M NaCl, 0.05% NaN_3 , pH 7.5, 0.5% octyl β -D-glucoside) were adsorbed to polystyrene slides overnight at 4 °C. Substrates were washed with PBS and quenched with 2% human serum albumin (HSA). CD34-site density was determined by saturation binding of radiolabelled monoclonal antibody to CD34 (ref. 18). Site densities yielded by input concentrations of ≤ 12 ng ml^{-1} were extrapolated from higher site-density determinations; as the immobilization procedure and the concentration of the detergent in the adsorption media were constant, site density of adsorbed CD34 was assumed to be proportional to CD34 concentration.

Peripheral blood lymphocytes and neutrophils were purified as described²³. Cells were stored in Ca^{2+} - and Mg^{2+} -free Hank's balanced salt solution (HBSS) containing 10 mM HEPES, pH 7.5, and 0.25% HSA (H/H medium), and resuspended at 5×10^5 cells per ml in binding medium (HBSS/HEPES containing 2 mM Ca^{2+}) before use. CD34 substrates were assembled as the lower wall in the flow chamber and mounted on an inverted phase-contrast microscope¹⁸. Images from a Nikon plan $\times 20$ objective were recorded on videotape. For detachment and rolling-velocity assays²³, cells were perfused through the chamber at 0.84 dyn per cm^2 until enough had accumulated (~ 3 min). Non-adherent cells were cleared by perfusion with binding medium at 0.84 dyn per cm^2 . Rolling velocities were determined for 30–50 of the cells observed during detachment assays. Cell displacement was measured over 5–8-s intervals. Velocities were measured only for cells that remained adherent throughout the 10-s period during which a given shear was applied. Cell accumulation was determined by perfusing cells (10^6 ml^{-1}) through the chamber at a range of wall shear stresses for 3 min and counting the rollingly adherent cells.

Treatments. Mild treatment with periodate was carried out after baseline cell-adhesion assays in the flow chamber and washing out of cells. Sodium periodate solution (5 mM) in PBS (pH 7.2) was infused into the flow chamber for 30 min at 4 °C in the dark¹⁶. After further cell-adhesion assays, in many cases substrates were reduced by infusion with 100 mM sodium borohydride (Sigma Chemical) in PBS pH 7.2, at room temperature, for 30 min. After each treatment, the flow chamber was equilibrated with binding medium and adhesion was scored to the same microscopic field of CD34. Cells were detached with 5 mM EDTA in H/H medium before substrate treatments and between tethering and accumulation assays at different shear stresses.

Inhibition with anti-L-selectin or control (IgG) monoclonal antibodies and with fucoidan and EDTA was performed as described¹⁸. All monoclonal antibodies and inhibitors remained during the adhesion assay.

Transient tethers. The duration of transient tethers on CD34 was analysed by a computerized imaging system, consisting of a Pentium computer with MVC 150/40-VL boards (Imaging Technology)²⁴. Cells interacting with the substrate were detected by discriminating their velocity from the hydrodynamic velocity

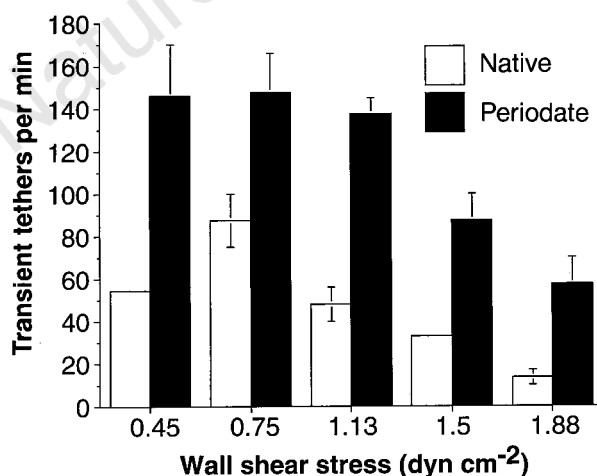


Figure 4 Mild treatment of CD34 with periodate enhances the transient-tethering frequency of neutrophils. We determined the frequency of transient tethers to CD34 at 0.5 sites per μm^2 before and after mild periodate treatment. Neutrophils were at a concentration of 10^6 ml^{-1} . Tethering events were defined as transient when no rolling ($<2 \mu\text{m}$ displacement) occurred while the cell was tethered. After a cell had transiently tethered once, it seemed to have a higher chance of making further transient tethers downstream. Therefore, for cells that made multiple tethers, only the first tether was counted.

distribution of cells free in flow. We analysed enough videotape to obtain 100 to 1,600 tethering events, and plotted the natural log of the number of cells that remained bound as a function of time after initiation of tethering. The most rapidly dissociating 90% or more of tethered cells were used to determine k_{off} .

Received 23 June 1997; accepted 13 February 1998.

1. Bell, G. I. Models for the specific adhesion of cells to cells: a theoretical framework for adhesion mediated by reversible bonds between cell surface molecules. *Science* **200**, 618–627 (1978).
2. Dembo, M., Torney, D. C., Saxman, K. & Hammer, D. The reaction-limited kinetics of membrane-to-surface adhesion and detachment. *Proc. R. Soc. Lond. B* **234**, 55–83 (1988).
3. Leckband, D., Müller, W., Schmitt, F.-J. & Ringsdorf, H. Molecular mechanisms determining the strength of receptor-mediated intermembrane adhesion. *Biophys. J.* **69**, 1162–1169 (1995).
4. Bruinsma, R. Les liaisons dangereuses: adhesion molecules do it statistically. *Proc. Natl Acad. Sci. USA* **94**, 375–376 (1997).
5. Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72**, 1541–1555 (1997).
6. Finger, E. B. *et al.* Adhesion through L-selectin requires a threshold hydrodynamic shear. *Nature* **379**, 266–269 (1996).
7. Alon, R., Fuhlbrigge, R. C., Finger, E. B. & Springer, T. A. Interactions through L-selectin between leukocytes and adherent leukocytes nucleate rolling adhesions on selectins and VCAM-1 in shear flow. *J. Cell Biol.* **135**, 849–865 (1996).
8. Lawrence, M. B., Kansas, G. S., Kunkel, E. J. & Ley, K. Threshold levels of fluid shear promote leukocyte adhesion through selectins (CD62L, P.E.). *J. Cell Biol.* **136**, 717–727 (1997).
9. Taylor, A. D., Neelamegham, S., Hellums, J. D., Smith, C. W. & Simon, S. I. Molecular dynamics of the transition from L-selectin to β_2 -integrin-dependent neutrophil adhesion under defined hydrodynamic shear. *Biophys. J.* **71**, 3488–3500 (1996).
10. Lasky, L. A. & Rosen, S. D. In *Inflammation: Basic Principles and Clinical Correlates* Vol. 2 (eds Gallin, J. I., Goldstein, I. J. & Snyderman, R.) 407–439 (Raven, New York, 1992).
11. Kansas, G. S. Selectins and their ligands: current concepts and controversies. *Blood* **88**, 3259–3287 (1996).
12. Springer, T. A. Traffic signals on endothelium for lymphocyte recirculation and leukocyte emigration. *Annu. Rev. Physiol.* **57**, 827–872 (1995).
13. Alon, R., Chen, S., Puri, K. D., Finger, E. B. & Springer, T. A. The kinetics of L-selectin tethers and the mechanics of selectin-mediated rolling. *J. Cell Biol.* **138**, 1169–1180 (1997).
14. Alon, R., Hammer, D. A. & Springer, T. A. Lifetime of the P-selectin: carbohydrate bond and its response to tensile force in hydrodynamic flow. *Nature* **374**, 539 (1995).
15. Norgard, K. E. *et al.* Enhanced interaction of L-selectin with the high endothelial venule ligand via selectively oxidized sialic acids. *Proc. Natl Acad. Sci. USA* **90**, 1068–1072 (1993).
16. Puri, K. D. & Springer, T. A. A Schiff base with mildly oxidized carbohydrate ligands stabilizes L-selectin and not P-selectin or E-selectin rolling adhesions in shear flow. *J. Biol. Chem.* **271**, 5404–5413 (1996).
17. Baumhueter, S. *et al.* Binding of L-selectin to the vascular sialomucin, CD34. *Science* **262**, 436–438 (1993).
18. Puri, K. D., Finger, E. B., Gaudernack, G. & Springer, T. A. Sialomucin CD34 is the major L-selectin ligand in human tonsil high endothelial venules. *J. Cell Biol.* **131**, 261–270 (1995).
19. Nicholson, M. W., Barclay, A. N., Singer, M. S., Rosen, S. D. & van der Merwe, P. A. Affinity and kinetic analysis of L-selectin (CD62L) binding to GlyCAM-1. *J. Biol. Chem.* **273**, 763–770 (1998).
20. Pierres, A., Benoliel, A.-M. & Bongrand, P. Experimental study of the rate of bond formation between individual receptor-coated spheres and ligand-bearing surfaces. *J. Phys. III France* **6**, 807–824 (1996).
21. Kyte, J. *Structure in Protein Chemistry* (Garland, New York, 1995).
22. Dembo, M. in *Some Mathematical Problems in Biology: Lectures on Mathematics in the Life Sciences* Vol. 25 (eds Goldstein, B. & Wofsy, C.) 1–27 (American Mathematical Society, Providence, 1994).
23. Puri, K. D., Finger, E. B. & Springer, T. A. The faster kinetics of L-selectin than of E-selectin and P-selectin rolling at comparable binding strength. *J. Immunol.* **158**, 405–413 (1997).
24. Chen, S., Alon, R., Fuhlbrigge, R. C. & Springer, T. A. Rolling and transient tethering of leukocytes on antibodies reveal specializations of selectins. *Proc. Natl Acad. Sci. USA* **94**, 3172–3177 (1997).

Acknowledgements. We thank T. K. Kishimoto, E. Butcher and G. Gaudernack for reagents.

Correspondence and requests for materials should be addressed to T.A.S.

Calcium oscillations increase the efficiency and specificity of gene expression

Ricardo E. Dolmetsch*†, Keli Xu* & Richard S. Lewis

Department of Molecular and Cellular Physiology, Beckman Center for Molecular and Genetic Medicine, Stanford University School of Medicine, Stanford, California 94305, USA

* These authors contributed equally to the work.

Cytosolic calcium ($[Ca^{2+}]_i$) oscillations are a nearly universal mode of signalling in excitable and non-excitable cells^{1–4}. Although Ca^{2+} is known to mediate a diverse array of cell functions, it is not known whether oscillations contribute to the efficiency or specificity of signalling or are merely an inevitable

consequence of the feedback control of $[Ca^{2+}]_i$. We have developed a Ca^{2+} clamp technique to investigate the roles of oscillation amplitude and frequency in regulating gene expression driven by the proinflammatory transcription factors NF-AT, Oct/OAP and NF- κ B. Here we report that oscillations reduce the effective Ca^{2+} threshold for activating transcription factors, thereby increasing signal detection at low levels of stimulation. In addition, specificity is encoded by the oscillation frequency: rapid oscillations stimulate all three transcription factors, whereas infrequent oscillations activate only NF- κ B. The genes encoding the cytokines interleukin (IL)-2 and IL-8 are also frequency-sensitive in a way that reflects their degree of dependence on NF-AT versus NF- κ B. Our results provide direct evidence that $[Ca^{2+}]_i$ oscillations increase both the efficacy and the information content of Ca^{2+} signals that lead to gene expression and cell differentiation.

Oscillations in $[Ca^{2+}]_i$ may be advantageous for receptor-mediated signal transduction, for example by increasing the fidelity of low-level signalling, preventing desensitization, or increasing signalling specificity^{1–4}; however, it has been difficult to demonstrate these and other possible functions⁵ for two reasons. First, the amplitude and frequency of $[Ca^{2+}]_i$ oscillations triggered through surface receptors varies among cells and in single cells over time, resulting in a mixture of stimulus waveforms that complicates analysis. Second, surface receptors are often coupled to multiple signalling pathways, making it difficult to ascribe downstream effects to Ca^{2+} alone. We have therefore developed a 'calcium clamp' method for generating homogeneous and synchronous receptor-independent $[Ca^{2+}]_i$ oscillations in large populations of T lymphocytes⁶ (Fig. 1). Ca^{2+} signals leading to T-cell activation are normally generated by a cascade involving antigen binding to the T-cell antigen receptor (TCR), generation of the second messenger inositol 1,4,5-trisphosphate ($InsP_3$), release of Ca^{2+} from internal stores, and Ca^{2+} influx across the plasma membrane⁷. Here we bypass the TCR/ $InsP_3$ pathway by treating Jurkat T cells with thapsigargin, an inhibitor of endoplasmic reticulum Ca^{2+} -ATPases that depletes internal Ca^{2+} stores and irreversibly activates store-operated Ca^{2+} (CRAC) channels in the plasma membrane⁸. Application of Ca^{2+} to cells treated in this way elevates $[Ca^{2+}]_i$ owing to influx through CRAC channels, whereas removal of extracellular Ca^{2+} allows pumps in the plasma membrane to return $[Ca^{2+}]_i$ to baseline levels. Thus, by rapidly changing the concentration of extracellular Ca^{2+} , it is possible to generate $[Ca^{2+}]_i$ oscillations having a uniform frequency across cells and an amplitude that is relatively constant in each cell over time (s.d., 12.7%) and among cells in the population (s.d., 18.2%; $n = 256$). A further advantage is that this technique probably mimics naturally occurring subcellular gradients of $[Ca^{2+}]_i$, because $[Ca^{2+}]_i$ oscillations triggered through the TCR result from periodic activation of CRAC channels^{9,10}.

We investigated whether oscillations affect the efficiency with which Ca^{2+} signals are detected. NF-AT is a Ca^{2+} -dependent transcription factor expressed in many cells, including T lymphocytes, in which it helps to regulate several immune-response genes including IL-2, IL-4 and tumour-necrosis factor- α (TNF- α)^{11,12}. NF-AT is activated by Ca^{2+} -stimulated dephosphorylation and translocation of a cytoplasmic subunit, which binds to a nuclear subunit induced by protein kinase C or by stimulation of the MAP kinase pathway^{11,12}. We compared the activity of an NF-AT/*lacZ* reporter gene¹³ in Jurkat cells stimulated with an oscillatory or constant elevation of $[Ca^{2+}]_i$ for 3 hours in the presence of 50 nM phorbol-12,13-dibutyrate (PdBu). Oscillations were generated with a period of 100 s, which is similar to the period in intact cells stimulated through the TCR^{9,14}. Oscillation amplitude was adjusted to produce the same average $[Ca^{2+}]_i$ as in the constant- $[Ca^{2+}]_i$ control cells to determine whether the kinetic features of the oscillations confer any signalling advantage relative to a sustained $[Ca^{2+}]_i$ increase, independently of the amount of Ca^{2+} that enters a cell. Figure 2a shows constant and oscillatory $[Ca^{2+}]_i$ stimuli in one

† Present address: Division of Neuroscience, Children's Hospital, Harvard University, 300 Longwood Avenue, Boston, Massachusetts 02120, USA.

experiment in which the average $[Ca^{2+}]_i$ was ~ 225 nM; subsequent fluorescence-activated cell sorting (FACS) analysis indicated that 39% of the cells stimulated with $[Ca^{2+}]_i$ oscillations were *lacZ*⁺, whereas only 19% of the cells with constant $[Ca^{2+}]_i$ were *lacZ*⁺. This enhancement of NF-AT-dependent transcription was not evident at higher $[Ca^{2+}]_i$ levels. In cells stimulated to an average of ~ 350 nM Ca^{2+} , oscillations activated roughly the same amount of *lacZ* expression as constant $[Ca^{2+}]_i$ elevation (Fig. 2b). The results of 12 paired experiments show that the induction of *lacZ* by constant $[Ca^{2+}]_i$ elevation falls off sharply below ~ 300 nM $[Ca^{2+}]_i$, whereas induction by oscillations is approximately constant down to at least 200 nM (Fig. 2c). The enhancement of NF-AT-dependent gene expression by oscillations thus increases as average $[Ca^{2+}]_i$ decreases below ~ 300 nM (Fig. 2d). In this way, oscillations appear to increase the ability of small Ca^{2+} signals to activate NFAT.

A long-standing question in cell signalling is how Ca^{2+} , with its abundant and varied intracellular targets, is able to achieve specificity and activate only a subset of those targets. One explanation is that variations in oscillation amplitude or frequency may discriminate among different Ca^{2+} -activated signalling pathways. In T cells, several transcription factors in addition to NF-AT are Ca^{2+} -sensitive, including NF- κ B and Oct/OAP^{15,16}. Both factors are involved in the control of IL-2 transcription, and NF- κ B also helps to regulate a large number of genes encoding cytokines, growth factors, adhesion molecules and other surface proteins^{11,17}. To determine whether the amplitude of the Ca^{2+} signal can distinguish among transcriptional pathways, we compared the Ca^{2+} sensitivity of luciferase reporter genes driven by NF-AT, Oct/OAP or NF- κ B in Jurkat cells. Phorbol ester (50 nM PdBu) or increased $[Ca^{2+}]_i$ (>600 nM) alone did not activate NF-AT or Oct/OAP measurably in thapsigargin-treated cells, and only slightly stimulated NF- κ B-dependent expression (~ 2.5 -fold above background). In contrast, the two stimuli acted together in a synergistic fashion to activate NF-AT, Oct/OAP and NF- κ B by 150-, 31- and 16-fold, respectively. In each case, the Ca^{2+} -dependent component of the response was completely suppressed by cyclosporin A, a potent and selective inhibitor of calcineurin,

consistent with a role for this phosphatase in mediating the effects of Ca^{2+} on these factors^{15,18,19}. As shown in Fig. 3a, all three pathways show a similar and highly nonlinear dependence on steady-state $[Ca^{2+}]_i$. The pronounced nonlinearity of NF-AT activation is consistent with previous studies^{5,20} and is an important determinant of the effects of oscillations on NF-AT signaling efficiency (see below). However, the similar Ca^{2+} dependencies of the three transcription factors suggests that oscillation amplitude is not likely to contribute significantly to selectivity among these pathways in T cells. In a previous study of B cells²¹, nuclear translocation of NF-AT appeared to be more Ca^{2+} -sensitive than that of NF- κ B; however, the transcriptional activity of these factors was not tested.

As oscillation amplitude is unlikely to discriminate among these three pathways, we tested whether oscillation frequency could. The calcium clamp was used to generate high-amplitude oscillations (~ 1 μ M peak) with periods from 100 to 1,800 s. As shown in Fig. 3b, the activation of all three transcription factors declines with increasing period. However, NF- κ B activation extends to oscillation periods as long as 1,800 s, in contrast to NF-AT and Oct/OAP, whose activity vanishes at periods ≥ 400 s. Frequency dependence of NF-AT has also been observed in response to photolytically generated pulses of $InsP_3$ (ref. 22), indicating that this behaviour is not specific to the use of thapsigargin to deplete stores. The difference in frequency requirements among the three factors is consistent with previous findings that NF- κ B persists in the nucleus for >16 min following a brief Ca^{2+} spike²¹, whereas NF-AT quickly exits following its rapid rephosphorylation^{21,23}. Thus, oscillation frequency dictates which combination of transcription factors is active; low frequencies recruit NF- κ B alone, whereas high frequencies activate NF-AT, Oct/OAP and NF- κ B. These results indicate that $[Ca^{2+}]_i$ oscillation frequency can discriminate among different transcriptional pathways.

The distinct frequency sensitivities of NF-AT, Oct/OAP and NF- κ B suggest that endogenous genes under the control of these factors may also respond differentially to oscillation frequency. We therefore compared the frequency sensitivities of genes encoding IL-2, an

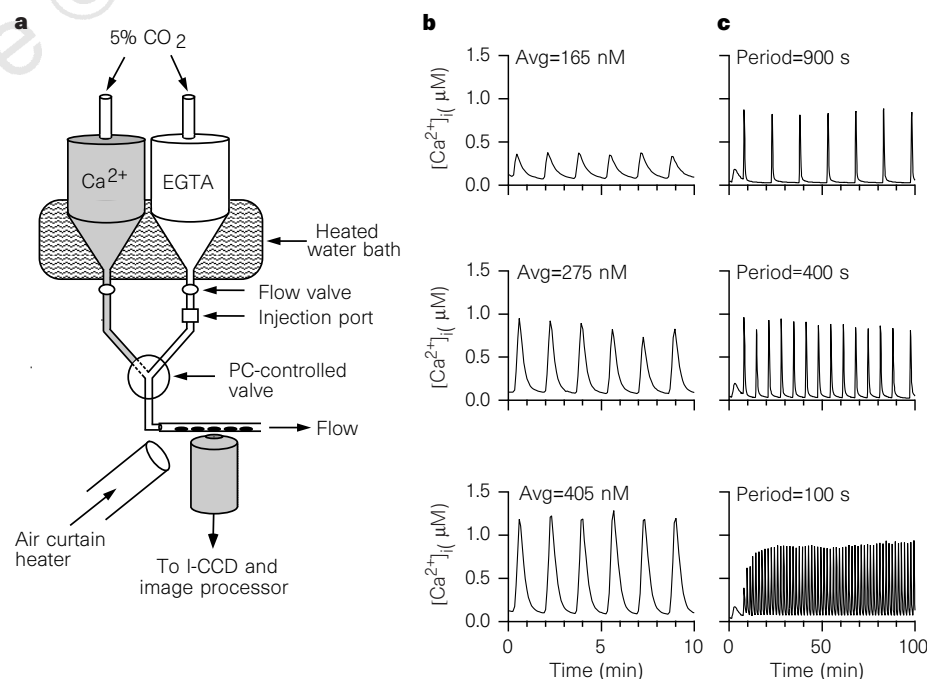


Figure 1 Generation of synchronized $[Ca^{2+}]_i$ oscillations in T cell populations. **a**, Diagram of the Ca^{2+} clamp (see Methods). Oscillations are generated by alternately exposing cells with depleted stores to 0 or 1.5 mM Ca^{2+} . **b**, **c**, Average $[Ca^{2+}]_i$ in 200–300 cells in which the amplitude or frequency of oscillations was

varied. In **b**, Ca^{2+} was applied for 10 s (top), 15 s (middle), or 30 s (bottom) every 100 s. In **c**, Ca^{2+} was applied for 30-s periods with a interspike interval of 900 s (top), 400 s (middle), or 100 s (bottom).

essential growth factor for T cells, and IL-8, a chemokine involved in recruiting immune cells to sites of infection. Both genes can be activated in T cells by Ca^{2+} ionophore and phorbol ester^{24,25}, but their promoters are driven by different combinations of transcription factors: IL-2 expression is highly dependent on NF-AT and Oct/OAP binding^{11,26}, whereas IL-8 depends on NF- κ B and other factors but not on NF-AT^{25,27}. Cells were transfected with luciferase reporter constructs driven by the IL-2 or IL-8 promoters and were stimulated for 3 h with 50 nM PdBu and constant $[\text{Ca}^{2+}]_i$ elevation. Both genes appear to be highly Ca^{2+} -sensitive, with IL-8 being more sensitive than IL-2 (Fig. 3c). The two genes show different sensitivities to $[\text{Ca}^{2+}]_i$ oscillation frequency (Fig. 3d): IL-2 is not expressed at periods ≥ 400 s, whereas IL-8 retains $\sim 20\%$ of its maximal activity under these conditions. The behaviour of IL-2 resembles that of NF-AT and Oct/OAP, whereas the behaviour of IL-8 is similar to that of NF- κ B. These results were supported by measurements of endogenous IL-8 expression in Jurkat cells treated with monensin to prevent secretion²⁸ (IL-2 expression after 3 h stimulation was too low to measure accurately). FACS analysis of anti-IL-8-stained cells indicated that the frequency dependence of IL-8 expression was similar to that of the reporter gene. Together, these results indicate

that oscillations of different frequencies can lead to the expression of different sets of genes, presumably as a consequence of their effects on the underlying transcription factors.

Our results demonstrate two important properties of nuclear signalling by $[\text{Ca}^{2+}]_i$ oscillations. First, oscillations enhance signalling efficiency specifically at low levels of stimulation. This effect arises from the highly nonlinear dependence of transcription on $[\text{Ca}^{2+}]_i$, so that oscillations periodically exceed the threshold for activation whereas a small constant $[\text{Ca}^{2+}]_i$ rise of the same average magnitude does not. The tendency of $[\text{Ca}^{2+}]_i$ to oscillate at low receptor occupancy in many cells may thus optimize sensitivity to weak external stimuli. Second, oscillations confer specificity on an otherwise highly pleiotropic Ca^{2+} signal. By differentially controlling the activation of distinct sets of transcription factors and the expression of different genes, oscillation frequency may direct cells along specific developmental pathways^{29,30}. Frequency-dependent gene expression is likely to be widespread, as Ca^{2+} -dependent transcription factors like NF- κ B and NF-AT are present in a great variety of cells and oscillations can occur with periods of seconds to hours^{1,29}. □

Methods

Cells and solutions. All experiments were done on the NF-ATZ-DIPA clone of Jurkat T cells, stably transfected with an NF-AT-*lacZ* reporter gene and selected for high *lacZ* inducibility¹³. Experiments were done in RPMI 1640 without

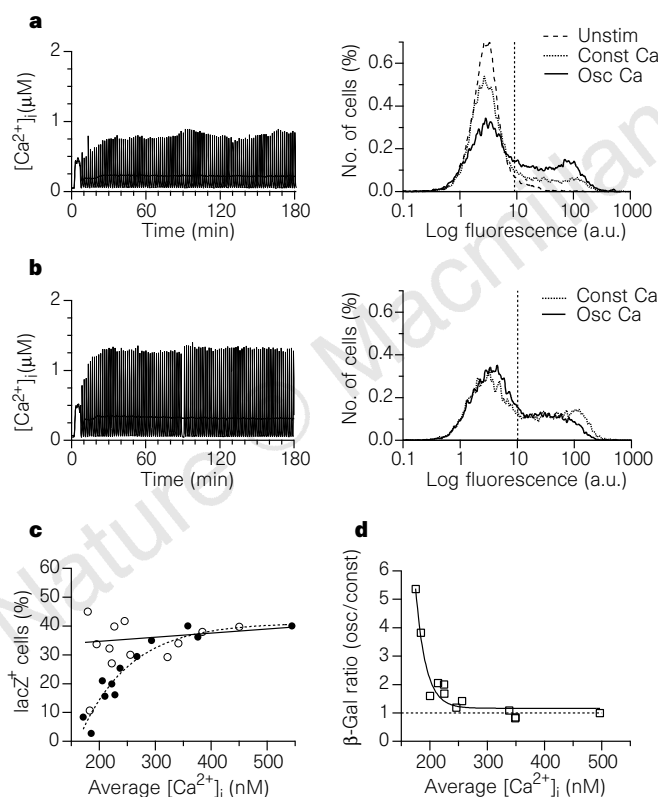


Figure 2 $[\text{Ca}^{2+}]_i$ oscillations enhance the calcium sensitivity of NF-AT-dependent transcription at low levels of stimulation. **a**, NF-AT-dependent *lacZ* expression in cells stimulated with a low level of $[\text{Ca}^{2+}]_i$. Left, superimposed $[\text{Ca}^{2+}]_i$ recordings from two cell populations (200–300 cells each) stimulated for 3 h with oscillating $[\text{Ca}^{2+}]_i$ (227 nM average) or constant $[\text{Ca}^{2+}]_i$ (225 nM average). Right, flow cytometry analysis of β -gal activity in the same experiments show that oscillating $[\text{Ca}^{2+}]_i$ activates 39% of the cells while constant $[\text{Ca}^{2+}]_i$ activates only 19%. The vertical gate excludes 97% of the unstimulated cells. **b**, NF-AT-dependent expression at a high level of $[\text{Ca}^{2+}]_i$. Left, stimulation with oscillating $[\text{Ca}^{2+}]_i$ (340 nM average) or constant $[\text{Ca}^{2+}]_i$ (359 nM average). Right, oscillating and constant $[\text{Ca}^{2+}]_i$ activate 33 and 39% of the cells, respectively. **c**, Percentage of *lacZ*⁺ cells induced by various concentrations of oscillating (○) or constant (●) $[\text{Ca}^{2+}]_i$. **d**, Ratio of the number of β -gal⁺ cells activated by oscillating as compared with constant $[\text{Ca}^{2+}]_i$, as a function of average $[\text{Ca}^{2+}]_i$, calculated from the data in **c**. All experiments were done in the presence of 50 nM PdBu.

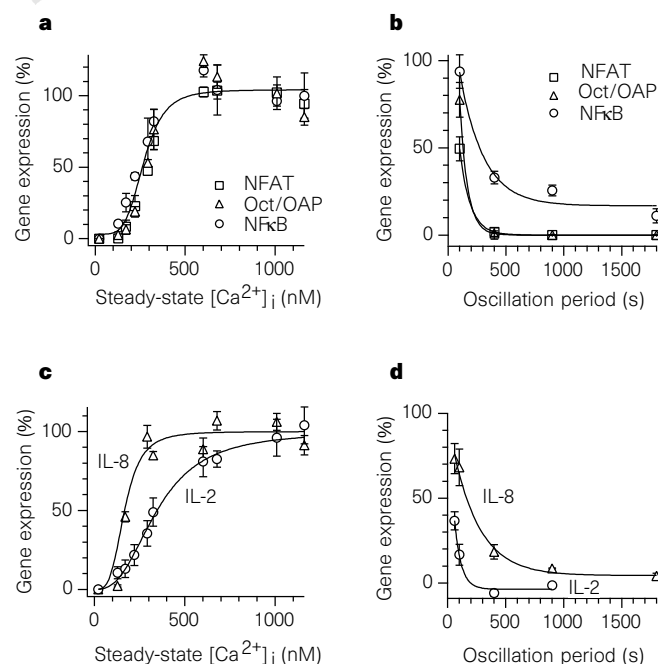


Figure 3 Effects of $[\text{Ca}^{2+}]_i$ and oscillation frequency on NF-AT-, Oct/OAP-, NF- κ B-, IL-2- and IL-8-luciferase reporter genes. **a**, Steady-state Ca^{2+} sensitivities of NF-AT, Oct/OAP and NF- κ B. A plot of the Hill equation is superimposed, with a Hill coefficient (n_H) of 4.7 and a K_d of 270 nM. **b**, Oscillation frequency dependences of NF-AT (□), Oct/OAP (Δ) or NF- κ B (○) reporter genes. **c**, Steady-state Ca^{2+} sensitivities of IL-2 (○) and IL-8 (Δ) reporter genes. Fits to the Hill equation are superimposed (for IL-2, $n_H = 2.7$ and $K_d = 346$ nM; for IL-8, $n_H = 3.6$ and $K_d = 166$ nM). **d**, Frequency dependences of IL-2 (○) and IL-8 (Δ) reporter genes. Error bars represent s.e.m. of triplicate measurements in 3–5 experiments; where they are not visible they are smaller than the symbols. All experiments were done in the presence of 50 nM PdBu.

phenol red or riboflavins and supplemented with 25 mM HEPES and 0.2% BSA. Solutions were maintained at 4 °C until immediately before each experiment, when they were warmed to 37 °C and supplemented with 50 nM PdBU and either 0.5 mM EGTA (~50 μ M free Ca^{2+}) or 1 mM Ca^{2+} (1.5 mM total Ca^{2+}).

Transfection of the reporter genes. NF-AT, NF- κ B, Oct-1/OAP and IL-2 reporter constructs were provided by J. Goldberg and G. Crabtree and were derived from constructs described previously¹⁶. Multimeric copies of NFAT (3 copies), NF- κ B (4) or Oct/Oap (4) binding sites, linked to a minimal (non-inducible) IL-2 promoter (-74 to +47), were inserted between the *Sma*I and *Hind*III sites in the multiple cloning region of the pGL-3 luciferase reporter vector (Promega). The construct with the intact IL-2 promoter region was made by inserting 371 bp of the IL-2 initiation region (-324 to +47) into the pGL-3 vector. The IL-8 reporter construct was provided by K. P. LeClair²⁷.

Transfection of 10^7 cells was by electroporation with 10 μ g of the reporter vector, 1 μ g of a vector containing large-T antigen and 2 μ g of a vector encoding the transmembrane and extracellular domains of CD8. Large-T antigen was used to increase the number of copies of the reporter construct in each cell, and the CD8 construct was used to determine the transfection efficiency. Transfection efficiencies were 30–40%; viability following centrifugation to remove cells killed during electroporation was 85–95%. Experiments were conducted 24–48 h after transfection when the expression of reporter genes was maximal.

Ca^{2+} clamp and Ca^{2+} measurements. Transfected cells were loaded with 2 μ M Fura-PE3/AM (Teflabs) for 1 h at 37 °C in loading medium (RPMI 1640, 25 mM HEPES, 2% fetal bovine serum), washed, and incubated for another hour to allow complete de-esterification of the dye. Loaded cells ($2-3 \times 10^5$) were allowed to adhere to a polylysine-coated laminar flow chamber and were placed on the heated stage of a Zeiss Axiovert 35 inverted microscope. The laminar flow chamber (60 μ l volume) was connected to two heated reservoirs containing 1.5 mM and 0 mM Ca^{2+} solutions, and pressurized with a mixture of 95% air and 5% CO_2 . At the start of each experiment, cells were treated with 1 μ M thapsigargin in 0 mM Ca^{2+} solution for 5 min to deplete internal Ca^{2+} stores and irreversibly activate CRAC channels. A computer-controlled solenoid valve (General Valve) was used to switch rapidly between the Ca^{2+} -containing and Ca^{2+} -free solutions flowing into the chamber and over the cells. The solution in the chamber was fully exchanged about once per second and was maintained at 37 °C. Cells were stimulated for 3 h while $[\text{Ca}^{2+}]_i$ was measured every 5–10 s by video microscopy as described¹⁰. Fura-PE3 was calibrated on the microscope in a microcuvette using solutions containing 1 mM EGTA and 10 mM Ca^{2+} according to ref. 10.

Reporter gene assays. Following stimulation cells were washed from the chamber, lysed by freeze/thawing and subjected to a luciferase assay using standard methods. All measurements were done in triplicate and normalized to the total number of cells determined using a Coulter Counter (Coulter Electronics). β -Galactosidase (β -gal) expression was measured by flow cytometry using the FACS-Gal protocol¹³. Steady-state $[\text{Ca}^{2+}]_i$ dependence of luciferase reporter genes (Fig. 3) was determined after 3 h stimulation with 1 μ M thapsigargin, 50 nM PdBU and variable extracellular $[\text{Ca}^{2+}]$. Luciferase activity was normalized to values between 0 and 100%, given by the responses to TG + PdBU in medium containing 0.5 mM EGTA or 1.5 mM Ca^{2+} , respectively.

Received 16 December 1997; accepted 12 February 1998.

- Berridge, M. J. & Galione, A. Cytosolic calcium oscillators. *FASEB J.* **2**, 3074–3082 (1988).
- Tsien, R. W. & Tsien, R. Y. Calcium channels, stores, and oscillations. *Annu. Rev. Cell Biol.* **6**, 715–760 (1990).
- Fewtrell, C. Ca^{2+} oscillations in non-excitable cells. *Annu. Rev. Physiol.* **55**, 427–454 (1993).
- Thomas, A. P., Bird, G. S. J., Hajnóczky, G., Robb-Gaspers, L. D. & Putney, J. W. Jr Spatial and temporal aspects of cellular calcium signaling. *FASEB J.* **10**, 1505–1517 (1996).
- Negulescu, P. A., Shastri, N. & Cahalan, M. D. Intracellular calcium dependence of gene expression in single T lymphocytes. *Proc. Natl Acad. Sci. USA* **91**, 2873–2877 (1994).
- Dolmetsch, R. E. & Lewis, R. S. in *Imaging Living Cells: A Laboratory Manual* (eds Konnerth, A., Lanni, F. & Yuste, R.) (CSHL Press, Cold Spring Harbor, in the press).
- Lewis, R. S. & Cahalan, M. D. Potassium and calcium channels in lymphocytes. *Annu. Rev. Immunol.* **13**, 623–653 (1995).
- Zweifach, A. & Lewis, R. S. Mitogen-regulated Ca^{2+} current of T lymphocytes is activated by depletion of intracellular Ca^{2+} stores. *Proc. Natl Acad. Sci. USA* **90**, 6295–6299 (1993).
- Lewis, R. S. & Cahalan, M. D. Mitogen-induced oscillations of cytosolic Ca^{2+} and transmembrane Ca^{2+} current in human leukemic T cells. *Cell Reg.* **1**, 99–112 (1989).
- Dolmetsch, R. & Lewis, R. S. Signaling between intracellular Ca^{2+} stores and depletion-activated Ca^{2+} channels generates $[\text{Ca}^{2+}]_i$ oscillations in T lymphocytes. *J. Gen. Physiol.* **103**, 365–388 (1994).

- Crabtree, G. R. & Clipstone, N. A. Signal transmission between the plasma membrane and nucleus of T lymphocytes. *Annu. Rev. Biochem.* **63**, 1045–1083 (1994).
- Rao, A. NF-ATp: a transcription factor required for the coordinate induction of several cytokine genes. *Immunol. Today* **15**, 274–281 (1994).
- Fiering, S. et al. Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. *Genes Dev.* **4**, 1823–1834 (1990).
- Donnadieu, E. et al. Imaging early steps of human T cell activation by antigen-presenting cells. *J. Immunol.* **148**, 2643–2653 (1992).
- Frantz, B. et al. Calcineurin acts in synergy with PMA to inactivate I κ B/MAD3, an inhibitor of NF- κ B. *EMBO J.* **13**, 861–870 (1994).
- Mattila, P. S. et al. The actions of cyclosporin A and FK506 suggest a novel step in the activation of T lymphocytes. *EMBO J.* **9**, 4425–4433 (1990).
- Baeuerle, P. A. & Henkel, T. Function and activation of NF- κ B in the immune system. *Annu. Rev. Immunol.* **12**, 141–179 (1994).
- Clipstone, N. A. & Crabtree, G. R. Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature* **357**, 695–697 (1992).
- Liu, J. et al. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **66**, 807–815 (1991).
- Fanger, C. M., Hoth, M., Crabtree, G. R. & Lewis, R. S. Characterization of T cell mutants with defects in capacitative calcium entry: Genetic evidence for the physiological roles of CRAC channels. *J. Cell Biol.* **131**, 655–667 (1995).
- Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C. & Healy, J. I. Differential activation of transcription factors induced by Ca^{2+} response amplitude and duration. *Nature* **386**, 855–858 (1997).
- Li, W.-h., Llopis, J., Whitney, M., Zlokarnik, G. & Tsien, R. Y. Cell-permeant caged InsP_3 ester shows that Ca^{2+} spike frequency can optimize gene expression. *Nature* **392**, 936–941 (1998).
- Shibasaki, F., Price, E., Milan, D. & McKeon, F. Role of kinases and the phosphatase calcineurin in the nuclear shuttling of transcription factor NF-AT4. *Nature* **382**, 370–373 (1996).
- Durand, D. B., Bush, M. R., Morgan, J. G., Weiss, A. & Crabtree, G. R. A 275 base pair fragment at the 5' end of the interleukin 2 gene enhances expression from a heterologous promoter in response to signals from the T cell antigen receptor. *J. Exp. Med.* **165**, 395–407 (1987).
- Okamoto, S. et al. The interleukin-8 AP-1 and κ B-like sites are genetic end targets of FK506-sensitive pathway accompanied by calcium mobilization. *J. Biol. Chem.* **269**, 8582–8589 (1994).
- Durand, D. B. et al. Characterization of antigen receptor response elements within the interleukin-2 enhancer. *Mol. Cell. Biol.* **8**, 1715–1724 (1988).
- Wechsler, A. S., Gordon, M. C., Dendorfer, U. & LeClair, K. P. Induction of IL-8 expression in T cells uses the CD28 costimulatory pathway. *J. Immunol.* **153**, 2515–2523 (1994).
- Prussin, C. & Metcalfe, D. D. Detection of intracytoplasmic cytokine using flow cytometry and directly conjugated anti-cytokine antibodies. *J. Immunol. Meth.* **188**, 117–128 (1995).
- Gu, X. & Spitzer, N. C. Distinct aspects of neuronal differentiation encoded by frequency of spontaneous Ca^{2+} transients. *Nature* **375**, 784–787 (1995).
- Fields, R. D., Eshete, F., Stevens, B. & Itoh, K. Action potential-dependent regulation of gene expression: temporal specificity in Ca^{2+} , cAMP-responsive element binding proteins, and mitogen-activated protein kinase signaling. *J. Neurosci.* **17**, 7252–7266 (1997).

Acknowledgements. We thank J. Goldberg for NF-AT-, Oct/OAP- and NF- κ B-luciferase constructs; K. LeClair and G. Crabtree for IL-8 and IL-2 luciferase constructs; C. Fanger for creating the large T antigen and CD8 constructs; G. Crabtree and L. Herzenberg for use of the luminometer and electroporator; S. Finkbeiner, Y. Gotoh and M. Greenberg for comments on the manuscript; and J. Walton for help in preparing constructs. This work was supported by a grant from the NIH (R.S.L.) and by a predoctoral fellowship from the American Heart Association, California Affiliate (R.E.D.).

Correspondence and requests for materials should be addressed to R.S.L. (e-mail: rslweis@leland.stanford.edu).

Cell-permeant caged InsP_3 ester shows that Ca^{2+} spike frequency can optimize gene expression

Wen-hong Li^{†‡}, Juan Llopis^{*}, Michael Whitney[§], Gregor Zlokarnik[§] & Roger Y. Tsien^{††}

Departments of ^{*} Pharmacology and [†] Chemistry & Biochemistry, and [‡] Howard Hughes Medical Institute, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0647, USA
[§] Aurora Biosciences Corporation, 11010 Torreyana Road, San Diego, California 92121, USA

Inositol 1,4,5-trisphosphate (InsP_3) releases calcium from intracellular stores and triggers complex waves and oscillations in levels of cytosolic free calcium^{1–5}. To determine which longer-term responses are controlled by oscillations in InsP_3 and cytosolic free calcium, it would be useful to deliver exogenous InsP_3 , under spatial and temporal control, into populations of unpermeabilized cells. Here we report the 15-step synthesis of a

[‡] Present address: Division of Biology 139-74, Beckman Institute, California Institute of Technology, Pasadena, California 91125, USA.

membrane-permeant, caged InsP_3 derivative from *myo*-inositol. This derivative diffused into intact cells and was hydrolysed to produce a caged, metabolically stable InsP_3 derivative. This latter derivative accumulated in the cytosol at concentrations of hundreds of micromolar, without activating the InsP_3 receptor. Ultraviolet illumination uncaged an InsP_3 analogue nearly as potent as real InsP_3 , and generated spikes of cytosolic free calcium, and stimulated gene expression via the nuclear factor of activated T cells^{6,7}. The same total amount of InsP_3 analogue elicited much more gene expression when released by repetitive flashes at 1-minute intervals than when released at 0.5- or ≥ 2 -minute intervals, as a single pulse, or as a slow sustained plateau. Thus, oscillations in cytosolic free calcium levels at roughly physiological rates maximize gene expression for a given amount of InsP_3 .

The physiological functions of oscillations in cytosolic free calcium ($[\text{Ca}^{2+}]_c$) levels and the existence of InsP_3 oscillations remain controversial¹⁻⁴. Stimulation with natural agonists does not allow dissection of the relative contributions of oscillation amplitude, frequency, number of pulses, and branching transduction pathways involving, for example, receptor tyrosine kinase activity or diacylglycerol production. It would be preferable to impose oscillations of various frequencies and a steady non-oscillatory elevation, and to compare their efficacies in triggering downstream events such as secretion or gene activation. One could drive $[\text{Ca}^{2+}]_c$ signals with either InsP_3 or extracellular Ca^{2+} ; the latter strategy⁸ also requires elevating the permeability of the plasma membrane to Ca^{2+} . We chose to use InsP_3 pulses because they mimic the natural cycle of calcium release from internal stores, work

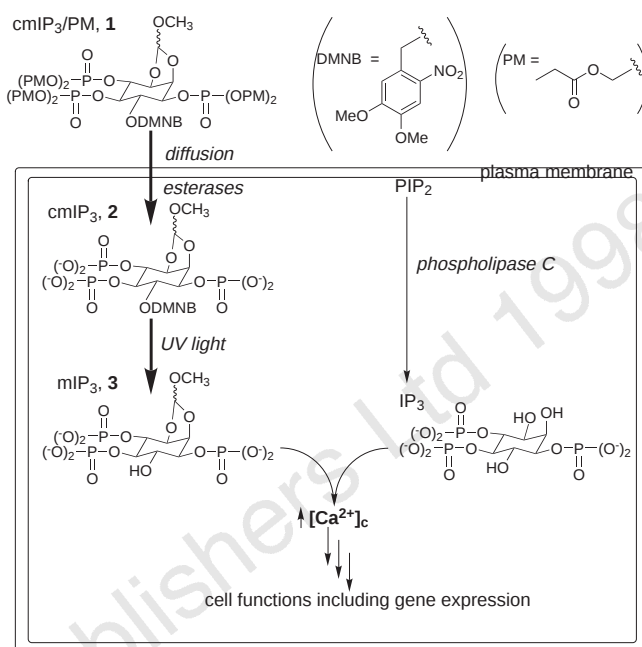


Figure 1 Mode of action of the caged, membrane-permeant InsP_3 derivative $\text{cmInsP}_3/\text{PM}$ (cmIP_3/PM , compound **1**), the trapped caged molecule (cmIP_3 , compound **2**), and 2,3-methoxymethylene- InsP_3 (mIP_3 , compound **3**) released by photolysis. PM, propionyloxymethyl group; DMNB, 4,5-dimethoxy-2-nitrobenzyl group.

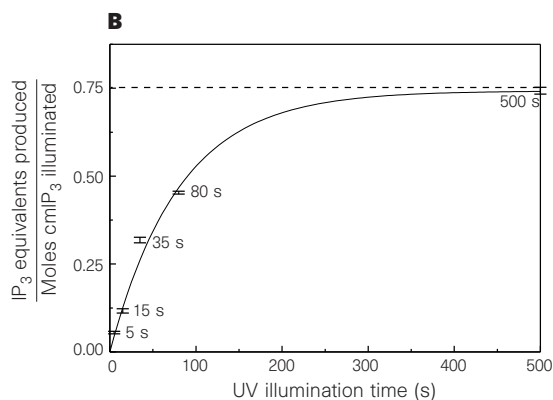
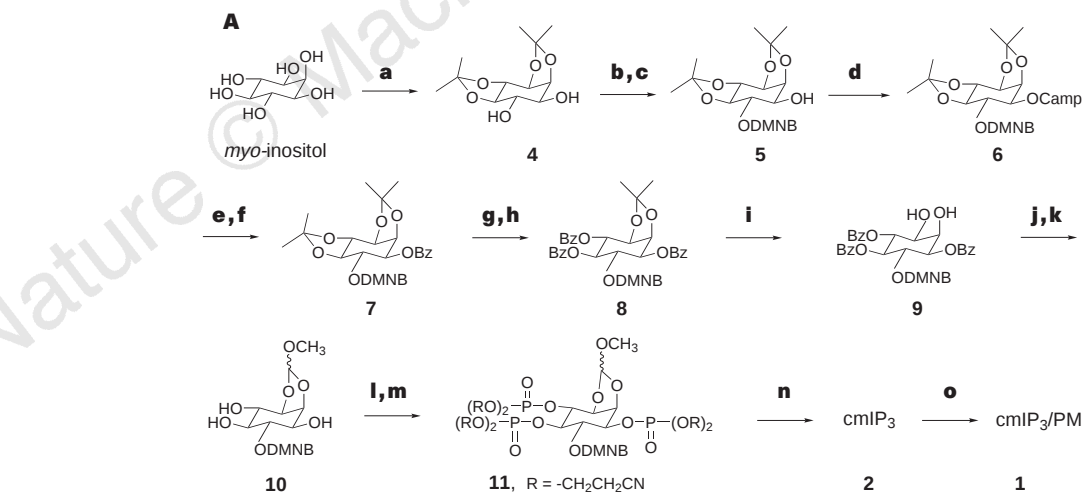


Figure 2 Synthesis and biochemical characterization of $\text{cmInsP}_3/\text{PM}$ and its reaction products. **A**, Chiral synthesis of $\text{cmInsP}_3/\text{PM}$, starting from *myo*-inositol. Bz, benzoyl; Camp, *S*-camphanyl. Reagents: **a**, 2-methoxypropene in acidic dimethylformamide (DMF), followed by chromatography and crystallization²⁶; **b**, Bu_2SnO , toluene azeotropy, 63% yield; **c**, DMNB bromide, CsF, DMF; **d**, *S*-(–)-camphanic acid chloride, triethylamine, 4-dimethylaminopyridine, then silica gel chromatography and crystallization to separate the diastereomeric camphanates, 42% of the desired isomer; **e**, K_2CO_3 , methanol/ CH_2Cl_2 ; **f**, BzCl, pyridine, 80%; **g**, $\text{HSCH}_2\text{CH}_2\text{OH}$, $\text{BF}_3/\text{Et}_2\text{O}$; **h**, BzCl, pyridine, 76%; **i**, $\text{HSCH}_2\text{CH}_2\text{OH}$, $\text{BF}_3/\text{Et}_2\text{O}$, 80%; **j**, $\text{HC}(\text{OMe})_3$, $\text{BF}_3/\text{Et}_2\text{O}$, DMF; **k**, K_2CO_3 , methanol/ CH_2Cl_2 , 52%; **l**, $(\text{NCCH}_2\text{CH}_2\text{O})_2\text{PN}(\text{i-Pr})_2$, tetrazole; **m**, *t*-butyl hydroperoxide, CH_2Cl_2 , 79%; **n**, NH_3 , aqueous methanol; **o**, $\text{EtCOOCH}_2\text{Br}$, $(\text{i-Pr})_2\text{NEt}$, CH_3CN , 12% for steps **n** + **o**. **B**, Binding to the InsP_3R , measured as equivalent picomol of genuine InsP_3 (IP_3), as a function of duration of ultraviolet-light-mediated photolysis of 8-pmol aliquots of cmInsP_3 (cmIP_3) *in vitro*. The smooth curve and dashed line show the fitted exponential progress curve (see Methods) and its asymptote, respectively.

at normal extracellular calcium levels, incorporate the complex relationship between InsP_3 and $[\text{Ca}^{2+}]_c$ and may be useful for many other studies.

Standard techniques for delivering the highly charged InsP_3 molecule across cell membranes, such as membrane permeabilization, microinjection, or patch clamping, would not allow reliable measurement of subsequent complex events, such as gene expression, in enough cells. To achieve precise temporal control without disrupting the plasma membrane, we needed a membrane-permeant derivative of caged InsP_3 , such as $\text{cmInsP}_3/\text{PM}$ (compound **1** in Fig. 1). All three phosphates of this compound were esterified with propionyloxymethyl (PM) groups, which hydrolyse in the cytoplasm to produce cmInsP_3 (compound **2** in Fig. 1). Free 2- and 3-hydroxyl groups are not obligatory^{9–11} for binding to the InsP_3 receptor (InsP_3R), so we masked these groups with a methoxymethylene group, which is relatively small and should not interfere with binding to the InsP_3R . Blockage of the 3-hydroxyl should prevent the released InsP_3 analogue from being phosphorylated by InsP_3 3-kinase to inositol 1,3,4,5-tetrakisphosphate. The 6-hydroxyl of InsP_3 interacts specifically with the InsP_3R and is essential^{10,11} for release of calcium, yet is difficult to regenerate from a permeant ester, perhaps because the bulky flanking 1- and 5-phosphate groups hinder the hydrolysis¹². Therefore we protected the 6-hydroxyl with a photolabile caging group, a 4,5-dimethoxy-2-nitrobenzyl (DMNB) ether¹³. Ultraviolet-mediated photolysis of the DMNB group should be independent of steric shielding by the flanking phosphates and should abruptly switch on biological activity through formation of compound **3** (2,3-methoxymethylene- InsP_3 , or mInsP_3). Also, in the absence of ultraviolet light, the ether linkage should be more resistant to metabolism or hydrolysis than the phosphodiester linkage previously used¹⁴ to cage InsP_3 .

cmInsP_3 was synthesized from *myo*-inositol in 14 steps, then esterified to produce $\text{cmInsP}_3/\text{PM}$ (Fig. 2A). Binding of cmInsP_3 to the InsP_3R (assayed by displacement of ^3H -labelled InsP_3) was

negligible ($<0.5\%$ of binding of real InsP_3). However, ultraviolet illumination of cmInsP_3 progressively released material that bound tightly to the InsP_3R (Fig. 2B). The product of exhaustive photolysis bound to the InsP_3R equivalently to 75% of an equimolar amount of real InsP_3 . Assuming that mInsP_3 was formed at 100% chemical yield, it is 75% as potent as InsP_3 , making mInsP_3 the most potent synthetic InsP_3 analogue known⁵. The quantum yield of photolysis was 0.09 ± 0.013 . At 365 nm, the uncaging cross-section of cmInsP_3 , $450 \text{ M}^{-1} \text{ cm}^{-1}$, slightly exceeded the $325 \text{ M}^{-1} \text{ cm}^{-1}$ reported for an earlier caged InsP_3 , a $\text{P}^{(4(5))}$ -1-(2-nitrophenyl)ethyl ester¹⁴.

Incubation of 1321N1 astrocytoma cells with $2 \mu\text{M}$ extracellular $\text{cmInsP}_3/\text{PM}$ for 5, 30 and 60 min resulted in 33 ± 5 , 201 ± 9 and $260 \pm 10 \text{ pmol}$ (mean $\pm$ s.e.m.), respectively, of trapped cmInsP_3 , per million cells. The final values were equivalent to intracellular concentrations of hundreds of micromolar. The ability of cmInsP_3 to accumulate to a much higher concentration inside the cells than the concentration of $\text{cmInsP}_3/\text{PM}$ outside the cells confirmed that cmInsP_3 was being trapped by irreversible ester hydrolysis, and implied that unphotolysed cmInsP_3 must be quite stable inside living cells. The trapped cmInsP_3 had no effect on $[\text{Ca}^{2+}]_c$ even after prolonged incubation, as monitored¹⁵ by the calcium indicator Fluo-3. Subsequent brief exposure of the cells to ultraviolet light (345–355 nm) caused a sudden increase in $[\text{Ca}^{2+}]_c$ levels whether extracellular calcium was present or not (Fig. 3a,b), showing that calcium was released from internal stores. Ultraviolet illumination was ineffective if applied less than two minutes after application of $\text{cmInsP}_3/\text{PM}$, as would be expected from the time required for the ester to cross membranes and release cmInsP_3 . We routinely allowed ≥ 30 min to ensure complete de-esterification of $\text{cmInsP}_3/\text{PM}$ after washing away excess ester. Once cmInsP_3 was loaded into cells, it remained well trapped and non-metabolized. At room temperature, an additional incubation time of 8 h in the dark had almost no effect on the size of the subsequent flash-induced $[\text{Ca}^{2+}]_c$ transient. Incubation for 8 h at 37°C did weaken the response, but the amplitude could be restored by a roughly threefold increase in exposure to ultraviolet light, indicating that about one-third of the cmInsP_3 remained. Other control experiments with the same or more than ten times stronger ultraviolet light, in the absence or presence of $100 \mu\text{M}$ 4,5-dimethoxy-2-nitrobenzyl methyl ether as a control for DMNB photolysis byproducts gave no detectable rise in $[\text{Ca}^{2+}]_c$ levels, either in normal calcium-containing or in calcium-free medium, showing that the response was specific to cmInsP_3 . $\text{cmInsP}_3/\text{PM}$ behaved similarly in other cell lines, including HeLa cells, human embryonic kidney 293 cells, REF-52 fibroblasts, T84 colonic epithelial cells, RBL-2H3 rat basophilic leukaemia cells (see below), and P388D1 macrophage-like cells, so its biological efficacy seems to be fairly general, at least in cultured mammalian cells.

$[\text{Ca}^{2+}]_c$ signals are essential for the activation of lymphoid cells, a process involving gene expression, differentiation and proliferation⁶. Non-invasive loading of $\text{cmInsP}_3/\text{PM}$ into cells allowed assessment of the effect of different temporal patterns of InsP_3R activation on $[\text{Ca}^{2+}]_c$ signals and gene expression. We used activation of the NF-AT (nuclear factor of activated T cells) transcription factor complex in RBL-2H3 cells as a model system. Increased $[\text{Ca}^{2+}]_c$ stimulates calcineurin to dephosphorylate pre-existing cytosolic NF-AT proteins (NF-ATc), which then migrate into the nucleus, bind to nuclear proteins and NF-AT-response elements, and activate downstream genes. Although NF-AT was first characterized in lymphocytes as a crucial element in interleukin-2 gene expression⁶, it is widespread in lymphoid cells, including RBL cells⁷. We assayed NF-AT-driven gene expression at the level of single cells, using a stably transfected reporter construct in which a trimer of NF-AT-response elements drove expression of the bacterial enzyme β -lactamase. β -Lactamase activity can be assayed with a newly developed membrane-permeant fluorogenic substrate, which becomes trapped in individual live cells. In the absence of β -lactamase, the cells fluoresce at 520 nm, whereas expression of the

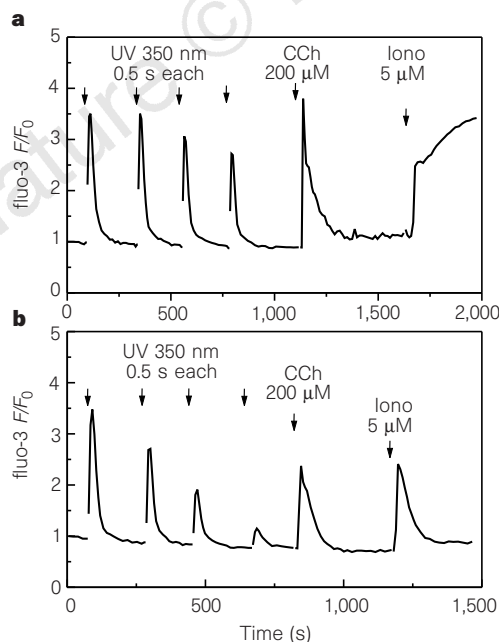


Figure 3 Cytosolic calcium transients evoked by photolysis of cmInsP_3 in 1321N1 astrocytoma cells. Transients were evoked in HBS medium containing normal levels of calcium (**a**) and in calcium-free DPBS saline containing 0.5 mM EGTA (**b**). UV 350 nm denotes uncaging flashes delivered through the microscope objective in which the excitation filter was switched for 0.5 s to a 345–355 nm bandpass. CCh, carbachol; Iono, ionomycin.

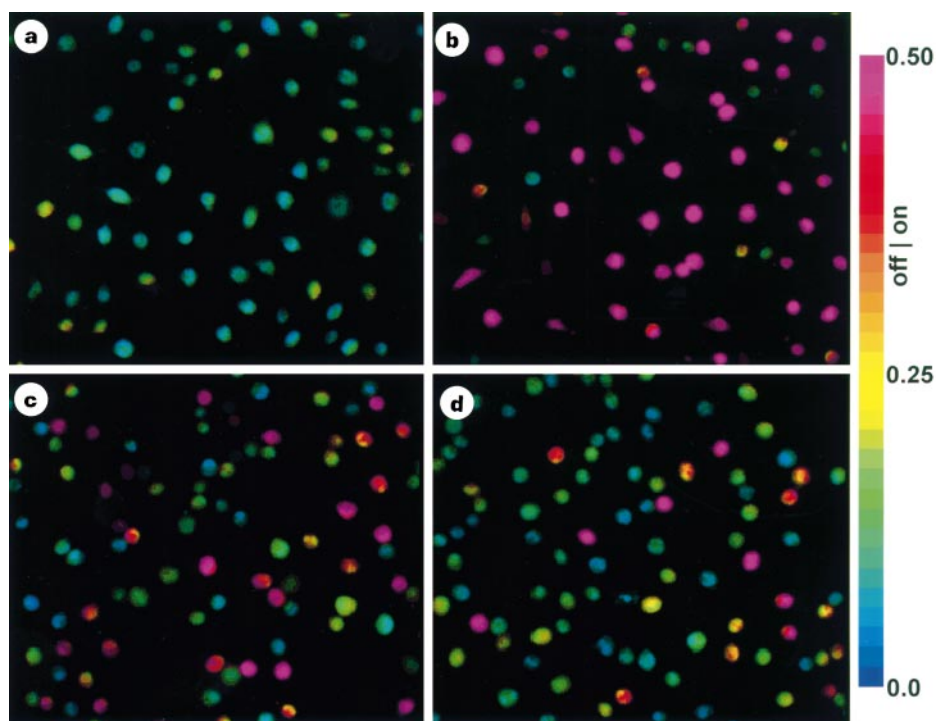


Figure 4 Expression of the β -lactamase reporter gene induced by uncaged cmlnsP₃/PM. **a**, Unstimulated cells. **b**, Cells stimulated by 1 μ M ionomycin for 20 min at 37 °C. **c**, **d**, Cells loaded with 10 μ M cmlnsP₃/PM were illuminated with four pulses each of ultraviolet light of 0.3, 0.6, 1 and 1.4 s duration, spaced 1 min (**c**) or 3 min (**d**) apart. The cells were then incubated for 3.5 h at 37 °C, and then stained

for β -lactamase gene expression¹⁶. The ratio of 420–460 nm to 512–558 nm emissions while exciting at 392–408 nm is displayed in false colours calibrated by the colour scale at the right, which also indicates the threshold ratio (0.35) for a cell to be scored as positive for reporter gene expression.

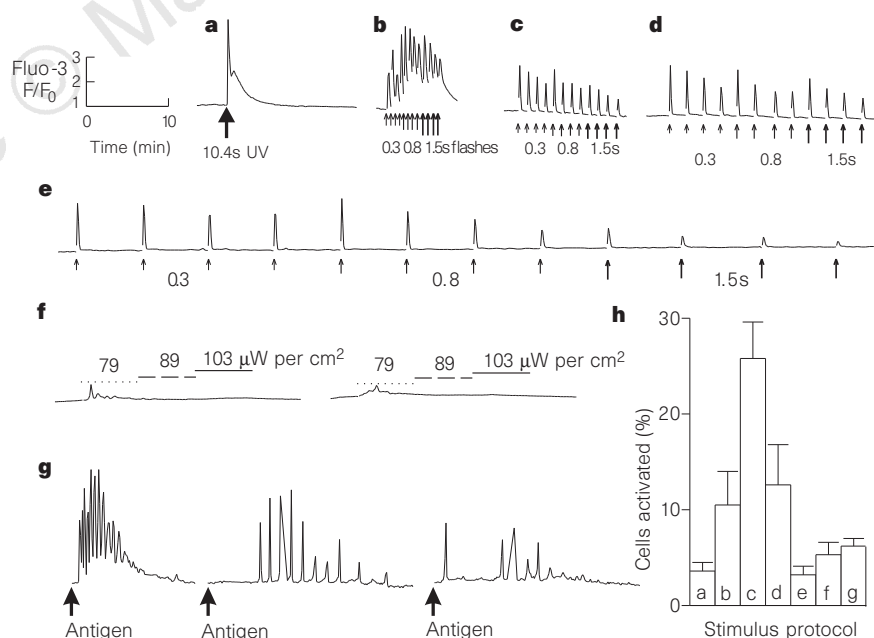


Figure 5 $[Ca^{2+}]_i$ levels measured by Fluo-3 fluorescence and NF-AT-driven β -lactamase gene expression evoked by different temporal patterns of photolysing the same total amount of cmlnsP₃ in RBL cells. **a–g**, The time and Fluo-3 intensity scales at the upper left apply to all traces. **a**, 3.5×10^{-8} einstein $\text{cm}^{-2} \text{s}^{-1}$ or 11.4 mW cm^{-2} was delivered for 10.4 s where indicated. **b–e**, 12 Pulses of the same intensity were delivered in three groups of four flashes, each of 0.3, 0.8, and 1.5 s duration, spaced 30 s (**b**), 1 min (**c**), 2 min (**d**), and 8 min (**e**) apart. **f**, Continuous illumination with (2.42, 2.7, or 3.1) $\times 10^{-10}$ einstein $\text{cm}^{-2} \text{s}^{-1}$, each for 7 min,

represented by dotted, dashed, and solid lines, respectively. **g**, 50 ng ml^{-1} DNP-BSA was added at the point indicated by the arrow. Traces in **f**, **g** are from representative single cells, whereas traces in **a–e** are averages of 25 cells because the responses were much more homogeneous and synchronous. **g**, The average frequency of spikes of amplitude $\Delta F/F_0 > 0.5$ was 2.09, 0.66, and 0.39 min^{-1} at the left, centre and right, respectively. **h**, NF-AT-driven β -lactamase gene expression. Percentage of cells is shown with suprathreshold β -lactamase activity resulting from protocols **a–g**.

enzyme cleaves the substrate and changes the fluorescence maximum to 447 nm¹⁶. As this new reporter gene assay leaves the cells fully viable, we could use fluorescence-activated cell sorting to isolate a subclone of RBL cells in which NF-AT-driven gene expression was maximally responsive to sustained elevation of $[Ca^{2+}]_c$ levels. Figure 4 shows pseudocolour images of β -lactamase expression levels in this subclone under various stimulation conditions. Hues from blue to magenta indicate low to high levels of NF-AT-driven expression of β -lactamase, monitored by ratios of emissions at 447 nm and 520 nm. The responsiveness of NF-AT activity to $[Ca^{2+}]_c$ levels was shown by the contrast between the blue-green pseudocolour of unstimulated cells (Fig. 4a) and the predominant magenta pseudocolour of cells treated with a calcium ionophore (Fig. 4b). Uncaging of cmInsP₃ with 16 ultraviolet pulses spaced 1 min apart caused significant gene activation (a 3.5-fold increase in ratio, corresponding to a pseudocolour of orange or redder) in ~30% of the cells (Fig. 4c), whereas the same flashes spaced 3 min apart only activated ~10% of the cells (Fig. 4d).

To confirm and extend this preliminary finding of frequency selectivity, we used a wider variety of uncaging protocols and made parallel measurements of $[Ca^{2+}]_c$ in the experiments shown in Fig. 5. For Fig. 5a, the ultraviolet light was applied in a single pulse of 11.4 mW per cm² that lasted 10.4 s, enough to photolyse ~30% of the loaded cmInsP₃. This generated a big $[Ca^{2+}]_c$ spike, which gradually returned to the resting level in <5 min. For Fig. 5b–e, the same intensity of ultraviolet light was applied in four pulses of 0.3 s, then four pulses of 0.8 s, and finally four pulses of 1.5 s duration; flashes were spaced 30 s (Fig. 5b), 1 min (Fig. 5c), 2 min (Fig. 5d), or 8 min (Fig. 5e) apart. Each flash produced a separate synchronized $[Ca^{2+}]_c$ spike; the increasing durations in the second and third groups of four pulses acted to maintain roughly constant spike amplitudes. In Fig. 5f, dim continuous illumination of 79, 89 and 103 μ W per cm² was applied for 7 min each, with no gaps. Each of the three 7-min episodes contained enough ultraviolet light to photolyse ~10% of the initial cmInsP₃. Even though the cumulative amount of mInsP₃ released in this protocol was similar to the amount of mInsP₃ released in the other protocols, and the overall duration of mInsP₃ release was intermediate between that of Fig. 5c and d, individual cells (for example, Fig. 5f) gave at most a few $[Ca^{2+}]_c$ transients, typically three to four, before subsiding. A roughly optimal dose of a relatively physiological agonist, surface immunoglobulin E crosslinked by dinitrophenylated albumin, evoked oscillations of extremely heterogeneous amplitude and frequency^{17,18} (Fig. 5g). Figure 5h summarizes the gene expression resulting from these different protocols. The most effective protocol used 12 flashes spaced 1 min apart; this activated NF-AT in 26% of the cells. This protocol was almost as effective as maintaining elevated $[Ca^{2+}]_c$ levels for 15 min with a calcium ionophore; this activated 28% of the cells. Spacings of 0.5 min and 2 min were each about half as effective as spacings of 1 min. The single 10.4 s pulse, the 12 pulses spaced 8 min apart, and the dim continuous illumination for 21 min all activated NF-AT in only 3–5% of cells. When either the cmInsP₃/PM or the ultraviolet illumination, or both, were withheld, negligible activation (<1%) resulted. Antigen stimulation was only modestly effective (6%), perhaps because the heterogeneous calcium responses matched optimal patterns in only a small fraction of the cells.

These results indicate that one of the best defined signal-transduction cascades into the nucleus may be tuned to the frequency of $[Ca^{2+}]_c$ spikes. The single burst of InsP₃ or excessively frequent oscillations of InsP₃ (Fig. 5a, b) may have failed to maintain elevated $[Ca^{2+}]_c$ levels for enough time^{19,20}. The lower-frequency oscillations (Fig. 5d, e) may have allowed too much time for rephosphorylation and nuclear exit of NF-ATc between pulses¹⁹. Slow, steady production of InsP₃ (Fig. 5f) was remarkably ineffective at increasing $[Ca^{2+}]_c$ levels for prolonged periods, perhaps because of InsP₃R desensitization^{21,22}. This observation may help to explain why cells

generate oscillations in $[Ca^{2+}]_c$ and perhaps InsP₃ as well^{2,18}. Cells might only be able to generate a limited total amount of InsP₃, because InsP₃ biosynthesis use many ATP molecules and depletes stores of the scarce lipid phosphatidylinositol-4,5-bisphosphate. Repetitive pulses (Fig. 5c) were much better than continuous dribbling of the same total amount of InsP₃ (Fig. 5f) for producing big and reliable $[Ca^{2+}]_c$ spikes, which in turn optimally activated at least one prototypic calcium-responsive transcription factor (see also ref. 8). Other genes²⁰ and cell functions²³ dependent on InsP₃ and $[Ca^{2+}]_c$ should also be investigated by this new, convenient method to activate the InsP₃R under spatiotemporal control in large populations of fully intact cells. □

Methods

Synthesis of cmInsP₃ and cmInsP₃/PM. See Fig. 2A. Fast atom bombardment mass spectroscopy showed an exact mass for compound 1 + Cs⁺ of 1,305.1481 compared with 1,305.1444 calculated for C₄₁H₆₂O₃₂NP₃ + Cs⁺. ¹H and ³¹P magnetic resonance spectra were also satisfactory.

Determination of the photolysis quantum yield of cmInsP₃ and affinity of binding of mInsP₃ to the InsP₃R. 1 μ M cmInsP₃ in Hanks' balanced salts (HBS) buffer plus 1 mg ml⁻¹ bovine serum albumin and 2 mM 2-mercaptoethanol was photolysed at 365 nm at 0 °C. Binding of mInsP₃ to the InsP₃R was measured (InsP₃ assay kit, Amersham) at the indicated times. The quantum yield, Q, of the photolysis was calculated from the exponential progress curve the extinction coefficient (5×10^6 cm² mol⁻¹) of cmInsP₃ at 365 nm, and the ultraviolet intensity (1.41×10^{-8} einstein·cm² s⁻¹) measured by ferrioxalate actinometry²⁴.

Determination of loading efficiency of cmInsP₃/PM. Confluent monolayers of astrocytoma cells in 35-mm tissue culture dishes were incubated with 2 μ M cmInsP₃/PM for the indicated time. After washing and 30 min incubation, the saline was aspirated and the cells were quenched with 0.5 ml ice-cold perchloric acid containing 2 mM EDTA and 20 mM HEPES. The cmInsP₃ was then thoroughly photolysed at 0 °C with 365 nm ultraviolet light. The samples were neutralized with KOH, centrifuged to remove KClO₄, and assayed for InsP₃ normalized to the number of cells. Experiments were done in duplicate on separate dishes of cells.

Monitoring cytosolic Ca²⁺ transients produced by uncaging cmInsP₃. 1321N1 astrocytoma cells were loaded with cmInsP₃/PM (2 μ M) and Fluo-3/AM (2 μ M) in HBS containing 0.05% Pluronic F-127 for 1 h at room temperature. Cells were then washed and incubated for another 30 min. Fluo-3 intensity¹⁵ (F , 480 $\pm$ 15 nm excitation, 535 $\pm$ 22.5 nm emission) was monitored every 7 s by digital-imaging microscopy and was plotted, after background subtraction, as a ratio against the Fluo-3 intensity at the resting calcium level (F_0). Ultraviolet pulses for the experiments of Fig. 3 were from a 150 W xenon arc, attenuated by a neutral density filter of 4% nominal transmission, filtered by a 345–355 nm bandpass filter, and delivered through a $\times 40/1.3$ NA oil-immersion objective. Typically, at least 13 cells were averaged, and the results shown are representative of three experiments. The same procedure was used for the RBL cells transfected with the reporter gene, except that the cmInsP₃/PM and Fluo-3/AM concentrations were 6 μ M and 1 μ M, respectively, 365 nm ultraviolet light was delivered to the entire dish of cells from a mercury lamp clamped above the stage of the inverted microscope, and the $[Ca^{2+}]_c$ responses of at least 25 cells were averaged. The short uncaging pulses of 11.4 mW per cm² or 3.5×10^{-8} einstein·cm⁻² s⁻¹ were gated by an electronic shutter. The $[Ca^{2+}]_c$ traces shown were obtained at room temperature. Responses at 37 °C to uncaging pulses were essentially the same as shown, and the asynchronous oscillations in response to the continuous illumination lasted a few more cycles but still terminated well before the end of the first 7 min of exposure to ultraviolet light.

Generation of an RBL cell line stably incorporating the NF-AT- β -lactamase transcriptional reporter. The pZeo-NF-AT-BLA vector was constructed by cloning a trimer of NF-AT-response-element promoters²⁵ upstream of a cytoplasmic form of β -lactamase¹⁶. This construct was made using the pZeoSV (Invitrogen) expression vector minus its SV40 promoter. RBL-2H3 cells were transfected by electroporation with pZeo-NF-AT-BLA DNA and selected in RPMI-1640 medium with 10% fetal bovine serum and 250 μ g ml⁻¹ zeocin for 2 weeks. The stably transfected population of RBL cells was

subcloned by stimulating with 1 μ M ionomycin for 3 h, vital staining with the membrane-permeant β -lactamase substrate CCF2/AM¹⁶, and fluorescence-activated cell sorting. The best clone gave 95–100% β -lactamase-positive cells with such stimulation and was used for all the subsequent studies on RBL cells.

Uncaging cmInsP₃ and measurement of reporter gene expression. The RBL cells were loaded with 10 μ M cmInsP₃/PM for 30 min (Fig. 4) or with 6 μ M ester for 1 h (Fig. 5h), and were then incubated in ester-free medium for 0.5 h at room temperature. Uncaging of cmInsP₃ was performed at 37 °C in culture medium in a sealed thermostated perfusion chamber (Biopetech, Butler, PA) with the same shuttered mercury lamp as for the [Ca²⁺]_i measurements. The cells were left for 3.5 h at 37 °C, loaded for 30 min at room temperature with 5 μ M of the β -lactamase substrate CCF2/AM in HBS containing 0.05% Pluronic F-127, and finally washed and incubated for 30 min at room temperature. β -Lactamase expression was scored by exciting at 392–408 nm wavelength and viewing the ratio of 420–460 to 512–558 nm emissions¹⁶. A charge-coupled-device camera (Photometrics) and ratio-imaging software (Metafluor, Universal Imaging) were used to obtain the data shown in Fig. 4 and to score borderline cells for Fig. 5h. Cells whose 420–460 nm to 512–558 nm ratio was >0.35, that is, 2.3 times that of control non-stimulated cells, were counted as activated. Over 200 cells were counted for each experiment, and the data shown in Fig. 5h are averages of three separate experiments. Over 1,200 cells were counted for each stimulation condition shown in Fig. 4.

Control stimuli for gene expression. For exposures to ultraviolet light in the absence of cmInsP₃/PM, we used the uncaging protocols of Fig. 5b or Fig. 5c. Ionomycin (1 μ M) treatments at 37 °C were terminated by washing with fresh medium containing 5 mg ml⁻¹ bovine serum albumin to sequester the ionophore. The cells were then incubated with 1.3 mM Ca²⁺ (Fig. 4b) or 3 mM EGTA. In the latter case, 15 and 30 min incubations with ionophore activated β -lactamase in 28 $\pm$ 5 and 65 $\pm$ 9% of the cells, respectively. To stimulate with antigen, the RBL cells were passively sensitized for 12 h with monoclonal murine IgE antibodies against dinitrophenyl (DNP) (Sigma). The cells were washed five times with fresh medium and incubated at 37 °C for 45 min before antigen stimulation. Dinitrophenylated bovine serum albumin (DNP-BSA, Calbiochem) was added at a final concentration of 50 or 500 ng ml⁻¹. 50 ng ml⁻¹ DNP-BSA caused the highest degree (in both amplitude and duration) of intracellular calcium mobilization, though the calcium transients in individual cells were heterogeneous. 500 ng ml⁻¹ DNP-BSA caused negligible mobilization of intracellular calcium and activated β -lactamase in only 2.1 $\pm$ 0.9% of cells.

Received 5 December 1997; accepted 6 March 1998.

- Berridge, M. J. Inositol triphosphate and calcium signalling. *Nature* **361**, 315–325 (1993).
- Tsien, R. W. & Tsien, R. Y. Calcium channels, stores, and oscillations. *Annu. Rev. Cell Biol.* **6**, 715–760 (1990).
- Jacob, R. Calcium oscillations in electrically nonexcitable cells. *Biochim. Biophys. Acta* **1052**, 427–438 (1990).
- Fewtrell, C. Ca²⁺ oscillations in non-excitable cells. *Annu. Rev. Physiol.* **55**, 427–454 (1993).
- Potter, B. V. L. & Lampe, D. Chemistry of inositol lipid mediated cellular signaling. *Angew. Chemie Int. Edn Engl.* **34**, 1933–1972 (1995).
- Crabtree, G. R. & Clipstone, N. A. Signal transmission between the plasma membrane and nucleus of T lymphocytes. *Annu. Rev. Biochem.* **63**, 1045–1083 (1994).
- Hutchinson, L. E. & McCloskey, M. A. Fc ϵ RI-mediated induction of nuclear factor of activated T-cells. *J. Biol. Chem.* **270**, 16333–16338 (1995).
- Dolmetsch, R. E., Xu, K. & Lewis, R. S. Calcium oscillations increase the efficacy and specificity of gene expression. *Nature* **392**, 933–936 (1998).
- Hirata, M. et al. Stereospecific recognition of inositol 1,4,5-trisphosphate analogs by the phosphatase, kinase, and binding proteins. *J. Biol. Chem.* **265**, 8404–8407 (1990).
- Hirata, M., Watanabe, Y., Yoshida, M., Koga, T. & Ozaki, S. Roles for hydroxyl groups of D-myo-inositol 1,4,5-trisphosphate in the recognition by its receptor and metabolic enzymes. *J. Biol. Chem.* **268**, 19260–19266 (1993).
- Kozikowski, A. P., Ognyanov, V. I., Fauq, A. H., Nahorski, S. R. & Wilcox, R. A. Synthesis of 1D-3-deoxy-, 1D-2,3-dideoxy-, and 1D-2,3,6-trideoxy-myo-inositol 1,4,5-trisphosphate from quebrachitol, their binding affinities, and calcium release activity. *J. Am. Chem. Soc.* **115**, 4429–4434 (1993).
- Li, W.-h., Schultz, C., Llopis, J. & Tsien, R. Y. Membrane-permeant esters of inositol polyphosphates, chemical synthesis and biological applications. *Tetrahedron* **53**, 12017–12040 (1997).
- Adams, S. R. & Tsien, R. Y. Controlling cell chemistry with caged compounds. *Annu. Rev. Physiol.* **55**, 755–784 (1993).
- Walker, J. W., Feeney, J. & Trentham, D. R. Photolabile precursors of inositol phosphates. Preparation and properties of 1-(2-nitrophenyl)ethyl esters of myo-inositol 1,4,5-trisphosphate. *Biochemistry* **28**, 3272–3280 (1989).
- Kao, J. P. Y., Harootyan, A. T. & Tsien, R. Y. Photochemically generated cytosolic calcium pulses and their detection by fluo-3. *J. Biol. Chem.* **264**, 8179–8184 (1989).
- Zlokarnik, G. et al. Quantitation of transcription and clonal selection of single living cells with β -lactamase as reporter. *Science* **279**, 84–88 (1998).

- Kuchty, J. & Fewtrell, C. Subcloning of RBL-2H3 mucosal mast cell line reduces Ca²⁺ response heterogeneity as the single-cell level. *J. Cell. Physiol.* **166**, 643–652 (1996).
- Smith, G. D., Lee, R. J., Oliver, J. M. & Keizer, J. Effect of Ca²⁺ influx on intracellular free Ca²⁺ responses in antigen-stimulated RBL-2H3 cells. *Am. J. Physiol.* **270**, C939–C952 (1996).
- Timmerman, L. A., Clipstone, N. A., Ho, S. N., Northrop, J. P. & Crabtree, G. R. Rapid shuttling of NF-AT in discrimination of Ca²⁺ signals and immunosuppression. *Nature* **383**, 837–840 (1996).
- Dolmetsch, R. E., Lewis, R. S., Goodnow, C. G. & Healy, J. I. Differential activation of transcription factors induced by Ca²⁺ response amplitude and duration. *Nature* **386**, 855–858 (1997).
- Hajnóczky, G. & Thomas, A. P. The inositol trisphosphate calcium channel is inactivated by inositol trisphosphate. *Nature* **370**, 474–477 (1994).
- Oancea, E. & Meyer, T. Reversible desensitization of inositol trisphosphate-induced calcium release provides a mechanism for repetitive calcium spikes. *J. Biol. Chem.* **271**, 17253–17260 (1996).
- Gu, X. & Spitzer, N. C. Distinct aspects of neuronal differentiation encoded by frequency of spontaneous Ca²⁺ transients. *Nature* **375**, 784–787 (1995).
- Hatchard, C. G. & Parker, C. A. A new sensitive chemical actinometer. II. Potassium ferrioxalate as a standard chemical actinometer. *Proc. R. Soc. Lond. A* **235**, 518–536 (1956).
- Fiering, S. et al. Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. *Genes Dev.* **4**, 1823–1834 (1990).
- Vacca, J. P. et al. The total synthesis of myo-inositol polyphosphates. *Tetrahedron* **45**, 5679–5702 (1989).

Acknowledgements. We thank E. A. Dennis for loan of facilities and S. R. Adams, A. Miyawaki, T. J. Rink and P. A. Negulescu for advice. This work was funded by the NIH and HHMI.

Correspondence and requests for materials should be addressed to R.Y.T.

NMR structure and mutagenesis of the FADD (Mort1) death-effector domain

Matthias Eberstadt, Baohua Huang, Zehan Chen, Robert P. Meadows, Shi-Chung Ng, Lixin Zheng*, Michael J. Lenardo* & Stephen W. Fesik

Pharmaceutical Discovery Division, Abbott Laboratories, Abbott Park, Illinois 60064, USA

*Laboratory of Immunology, National Institutes of Health, Bethesda, Maryland 20892, USA

When activated, membrane-bound receptors for Fas and tumour-necrosis factor initiate programmed cell death by recruiting the death domain of the adaptor protein FADD¹ (Mort1; ref. 2) to the membrane. FADD then activates caspase 8 (ref. 3) (also known as FLICE⁴ or MACH⁵) through an interaction between the death-effector domains of FADD and caspase 8. This ultimately leads to the apoptotic response. Death-effector domains and homologous protein modules known as caspase-recruitment domains⁶ have been found in several proteins^{1–14} and are important regulators of caspase (FLICE) activity and of apoptosis. Here we describe the solution structure of a soluble, biologically active mutant of the FADD death-effector domain. The structure consists of six anti-parallel, amphipathic α -helices and resembles the overall fold of the death domains of Fas¹⁵ and p75 (ref. 16). Despite this structural similarity, mutations that inhibit protein–protein interactions involving the Fas death domain have no effect when introduced into the FADD death-effector domain. Instead, a hydrophobic region of the FADD death-effector domain that is not present in the death domains is vital for binding to FLICE and for apoptotic activity.

As neither the wild-type death-effector domain (DED) of FADD nor full-length FADD could be obtained in a form that was suitable for study by nuclear magnetic resonance (NMR), we prepared several site-directed mutants and screened them for their ability to be expressed as a soluble protein, induce apoptosis and bind to FLICE. On the basis of its high solubility in a monomeric form (0.7 mM) and wild-type-like biological activities (Fig. 1), we chose the Phe 25 $\rightarrow$ Tyr (F25Y) mutant protein for structure determination.

The structure of the F25Y FADD DED is well defined by the 1,102 experimentally derived NMR restraints (Fig. 2a) and consists of six antiparallel, amphipathic α -helices (Fig. 2b). The α -helices are

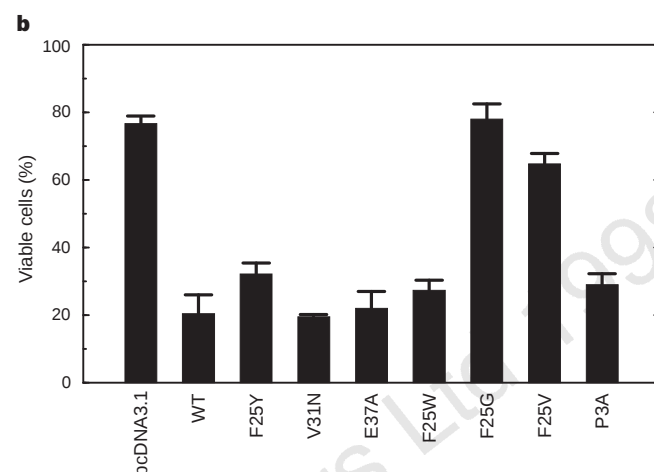
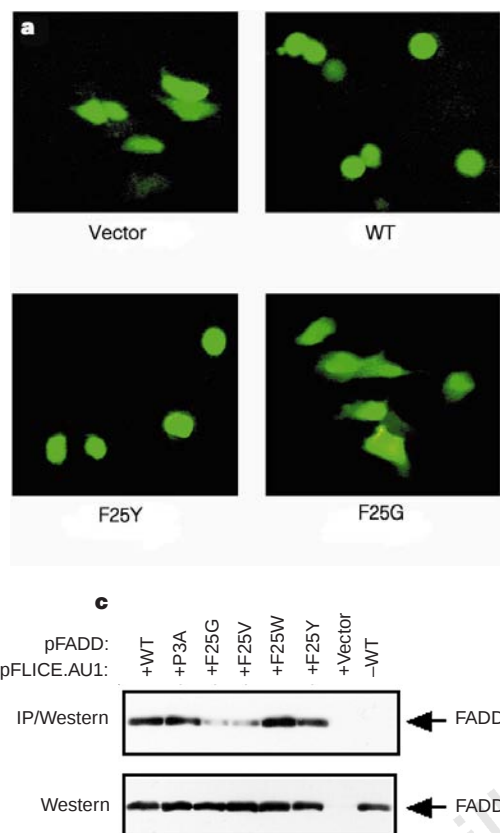


Figure 1 Biological activity of FADD death-effector-domain mutants. **a**, Functional effects of the expression of FADD and the DED mutants in T47D cells. Wild-type FADD and F25Y FADD induce apoptosis, as indicated by the characteristic rounded appearance of the cells. **b**, Quantification of FADD-induced apoptosis in T47D cells. pcDNA3.1, vector alone; WT, wild type. **c**, The association of FADD DED mutants with FLICE (caspase 8). Top, FADD proteins co-immunoprecipitated (IP) with FLICE; bottom, confirmation of FADD protein expression in cell lysates.

connected by short loops of irregular structure. The only exception is the loop connecting helices $\alpha 4$ and $\alpha 5$, which contains a β -turn. The surface of the death-effector domain contains a hydrophobic patch on each side of the protein. One site is composed of residue Y25 (F25 in the wild-type protein), L26, and L28 from $\alpha 2$ and L66 and L70 from $\alpha 5$ (Fig. 2c). The other site is composed of hydrophobic residues from $\alpha 1$ (M1, P3, L5 and V6), $\alpha 4$ (L43 and L50), $\alpha 6$, (F82) and the β -turn (P57) (Fig. 2d).

The overall fold of the FADD DED resembles the recently determined structures of the Fas and p75 death domains^{15,16}. The root-mean-square deviation (r.m.s.d.) for the superposition of 55 C α atoms from six α -helices of the FADD DED and p75 death domain is 2.1 Å. The Fas death domain compares less well to the FADD DED, with an r.m.s.d. of 3.2 Å. There are differences between the death-effector domain and the two death domains in the length and orientation of the helices, especially for $\alpha 3$, $\alpha 5$ and $\alpha 6$ (Fig. 3a, b). In addition, the loops vary in length among the three proteins. Furthermore, the FADD DED contains a β -turn between helices $\alpha 4$ and $\alpha 5$, which is a unique structural feature of this protein.

We have used site-directed mutagenesis to identify functionally important residues of the Fas death domain¹⁵. To probe the importance of analogous residues in the structurally similar FADD DED, we expressed mutant FADD proteins in a breast cancer cell line, T47D, and tested these mutants for their ability to induce apoptosis. In contrast to the lymphoproliferation (*lpr*) mutation of the Fas death domain, which alters the three-dimensional structure of the domain¹⁷ and blocks its activity¹⁸, the analogous mutation in the FADD DED (V31N) had no effect (Fig. 1b). Similarly, FADD DED with the E37A mutation, corresponding to another inactive mutant of the Fas death domain¹⁵, exhibited the same apoptotic activity as the wild-type protein. Thus, even though death domains and the FADD death-effector domain are structurally similar, the functionally important residues of these protein modules are distinct.

Unlike the Fas death domain, which is mostly charged on its surface, the FADD DED has two hydrophobic patches which could be important for its apoptotic activity. With the exception of L28, the hydrophobic residues from one of these sites are highly conserved among the death-effector domains of FLICE and the viral and cellular FLICE-inhibitory proteins (FLIPs) (Fig. 3a). F25, which is located in the middle of this hydrophobic site, can be substituted by tyrosine or tryptophan with little or no loss in apoptotic activity or in ability to bind to FLICE (Fig. 1). However, the F25G and F25V mutant proteins are inactive and exhibit weaker binding to FLICE compared with the wild-type protein (Fig. 1). As the loss of activity and the ability to bind to FLICE might be due to a structural alteration, we determined the three-dimensional structure of the F25G inactive mutant protein. The structure of this protein is well defined by 1,036 NMR-derived restraints and is similar to the active F25Y FADD DED (Fig. 4). Therefore, the hydrophobic region involving F25 is important for mediating protein-protein interactions. In contrast, the other hydrophobic region (Fig. 2d) may not be important for binding to FLICE and inducing apoptosis, as the hydrophobic residues that compose this site are not conserved (Fig. 3a). Furthermore, when a proline within this hydrophobic site (P3) was mutated to an alanine, there was no loss in activity or binding to FLICE (Fig. 1).

Death domains, death-effector domains, and caspase-recruitment domains are all involved in signalling programmed cell death and interact with protein modules within the same family^{1,2,4,5}. As shown here, the death-effector and death domains adopt similar three-dimensional structures but use different residues to induce apoptosis. The caspase-recruitment domains may adopt a similar structure to the death and death-effector domains, as previously proposed⁶, as the key hydrophobic residues that define the structural core of these latter two protein modules are also found in caspase-recruitment domains. However, the critical residues that are required for apoptotic activity in death-effector domains and in

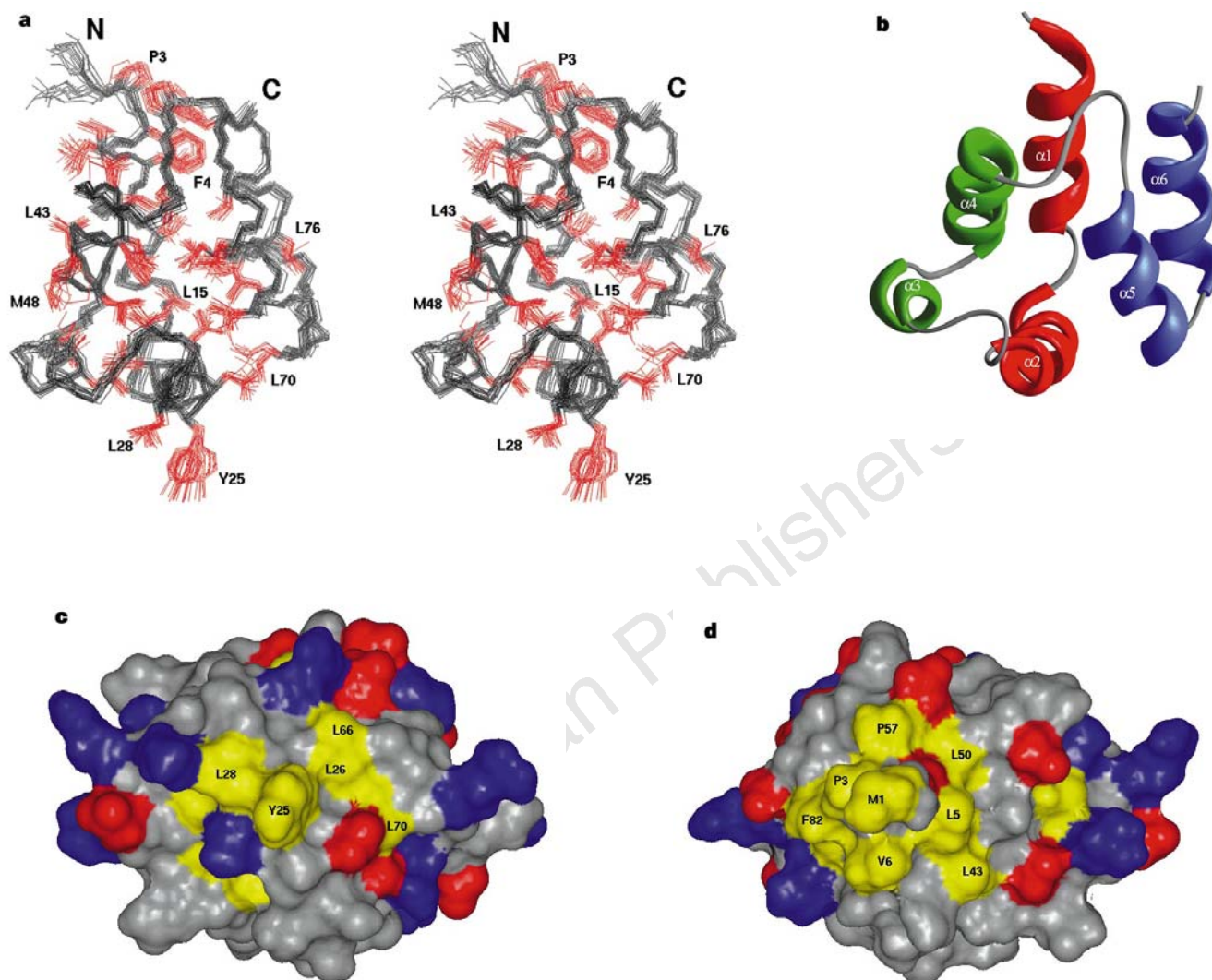


Figure 2 Structure of the F25Y mutant DED of FADD. **a**, Stereoview of the backbone (N, C α , C') of 20 superimposed NMR-derived structures of the F25Y FADD DED (residues 2–83). The side chains of hydrophobic residues are coloured red. The r.m.s.d. about the mean coordinate positions for residues 2–83 is $0.36 \text{ \AA} \pm 0.07 \text{ \AA}$ for the backbone atoms (N, C α , C') and $0.92 \text{ \AA} \pm 0.006 \text{ \AA}$ for all heavy atoms. No distance restraint is violated by more than $0.25 \text{ \AA}$ in any of the structures. For the ensemble of final structures, the XPLOR NOE potential energy is $0.15 \pm 0.05 \text{ kcal mol}^{-1}$. No torsional restraint is violated by more than 5° . For the

ensemble, the torsional energy is $0.35 \pm 0.02 \text{ kcal mol}^{-1} \text{ rad}^{-2}$. The non-bonded geometries are well satisfied as indicated by low van der Waals energies of $-520 \pm 9 \text{ kcal mol}^{-1}$. **b**, Ribbons²⁸ depiction of the averaged minimized NMR structure of the F25Y FADD DED. **c**, **d**, Two views of the surface representation of the averaged minimized NMR structure of the F25Y FADD DED. The side chains of hydrophobic (Met, Leu, Ile, Val, Pro, Phe, Tyr), basic (Lys, Arg), and acidic (Asp, Glu) residues are coloured yellow, blue, and red, respectively.

death domains^{15,19} are not conserved in caspase-recruitment domains, indicating that these domains may use a different set of residues to define their binding specificity and function. ☐

Methods

Apoptosis assay. The coding region of FADD and the mutants were cloned into the pcDNA3.1 expression vector (Invitrogen). A breast cancer cell line T47D (ATCC) expressing functional Fas²⁰ was used to assay the biological activities of the DED mutants of FADD. Transient transfection was carried out in the presence of $0.8 \mu\text{g}$ of the FADD-expression vector and equal molar amounts of the green-fluorescent-protein vector pEGFP-C3 (Clontech). After 20 h, more than 400 green cells were counted for each sample to assess the extent of cell death. We confirmed that the round green cells were apoptotic by propidium iodide staining of the nuclei¹.

The association of FADD DED mutants with FLICE. AU1-tagged FLICE was cloned into pCI^{neo} expression vector (Pharmacia); FADD was cloned into pcDNA3.1. The 293T cells were co-transfected with $0.5 \mu\text{g}$ FADD DNA

and $1 \mu\text{g}$ FLICE DNA. The cells were cultured for 20 h in the presence of a caspase inhibitor ($15 \mu\text{M}$ zVAD), lysed and immunoprecipitated with monoclonal antibody to AU1 (BabCO) and $20 \mu\text{l}$ protein A–Sepharose at 4°C for 2 h. We separated the co-immunoprecipitated proteins by SDS–PAGE and visualized them by western blotting using mouse anti-human FADD (Transduction Laboratories) and horseradish-peroxidase-conjugated goat anti-mouse IgG.

NMR sample preparation. FADD DED mutant proteins (F25Y and F25G) containing residues 1–83 and a polyhistidine tag (–LEHHHHHH) at the carboxy terminus were expressed in *Escherichia coli* HMS174(DE3) using the pET30b vector (Novagen). Site-directed mutagenesis was performed using the Quickchange kit (Stratagene). We prepared isotopically labelled proteins from cells grown on M9 medium containing $^{15}\text{NH}_4\text{Cl}$ and/or $[\text{U-}^{13}\text{C}]$ -glucose in either H_2O or $75\% \text{ } ^2\text{H}_2\text{O}$. We purified the FADD DED mutants by affinity chromatography on a nickel-IDA column (Invitrogen). NMR samples contained 0.7 mM protein in 50 mM sodium acetate (pH 4.0) and 50 mM NaCl in $\text{H}_2\text{O}/^2\text{H}_2\text{O}$ (9:1) or $^2\text{H}_2\text{O}$.

Figure 3 Comparison of death-effector and death domains. **a**, Alignment of DEDs (sequence-based) and two death domains (structure-based) with the FADD DED. Buried hydrophobic residues (yellow), residues contributing to the hydrophobic patch involving F25 (green), residues contributing to the second hydrophobic site in FADD DED (asterisks), and α -helices of the FADD DED are marked. Amino-acid sequences for DEDs are obtained from FADD¹², FLICE^{4,5}, c-FLIP¹¹⁻¹⁴, E8 and MC159 (refs 8-10). DD, death domain. **b**, Stereoview of the α -helices (represented by cylinders) of the superimposed averaged minimized NMR structures of the F25Y FADD DED (red), p75 death domain (blue) and Fas death domain (yellow).

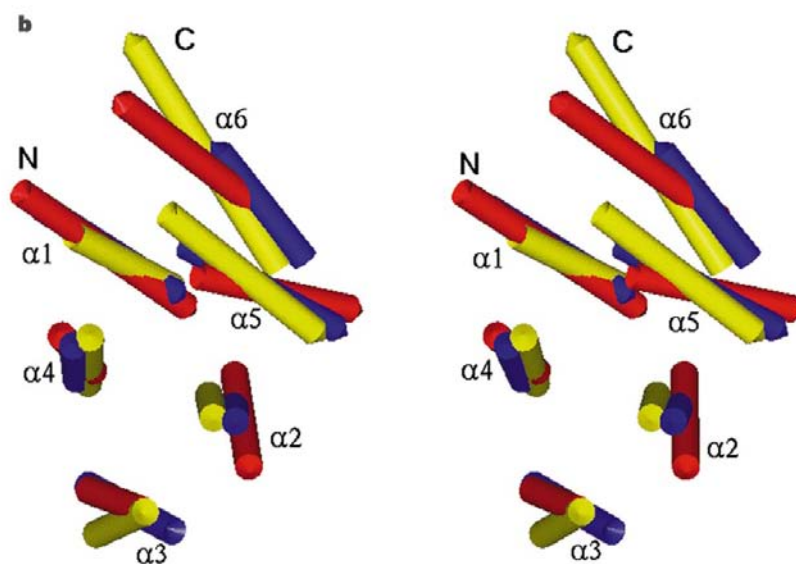
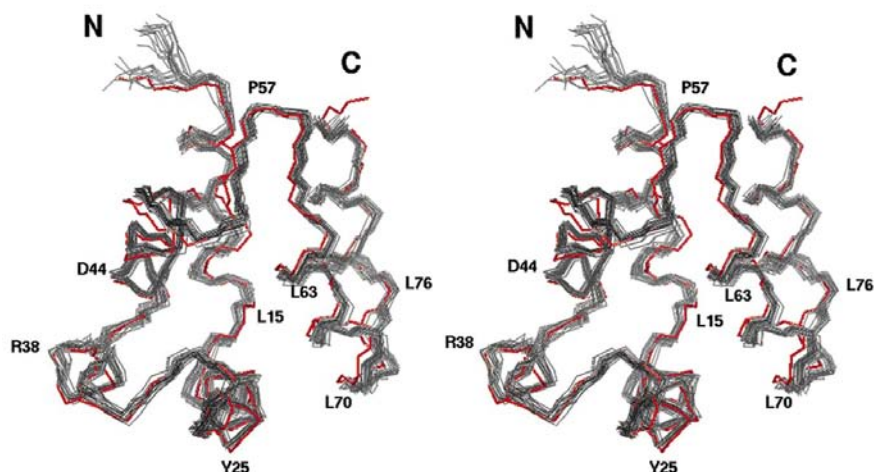


Figure 4 Stereoview of the backbone (N, C, C') of 20 superimposed NMR-derived structures of the inactive F25G FADD DED (black) overlaid onto the averaged NMR structure of the active F25Y FADD DED (red). The r.m.s.d. about the mean coordinate positions for residues 2-83 in the F25G mutant protein is $0.42 \pm 0.05 \text{ \AA}$ for the backbone atoms (N, C α , C') and $0.93 \pm 0.05 \text{ \AA}$ for all heavy atoms. The backbone r.m.s.d. between the F25G and F25Y FADD DEDs is $1.14 \text{ \AA}$ for residues 1-83.



Structure determination. All NMR spectra were acquired at 15 °C on Bruker NMR spectrometers. ^1H , ^{13}C and ^{15}N resonances of the backbone were assigned by deuterium-decoupled triple-resonance experiments²¹. ^{13}C and ^1H resonances of the side chain were assigned from three-dimensional HCCH-TOCSY²², CBCA(CO)NH²² and HC(CO)NH-TOCSY²³ experiments. Stereospecific assignment of methyl groups of the valine and leucine residues were obtained using a fractionally ^{13}C -labelled sample²⁴. We obtained distance restraints from ^{15}N - or ^{13}C -resolved three-dimensional NOESY experiments²². ϕ -angle restraints were based on the $^3J_{\text{HNH}\alpha}$ coupling constants measured in an HNHA experiment²⁵. Slow-exchanging amide protons were identified from a series of $^1\text{H}/^{15}\text{N}$ HSQC spectra recorded after the H_2O buffer was changed to a $^2\text{H}_2\text{O}$ buffer. Structures of the mutant FADD death-effector domains were calculated with a distance geometry/simulated annealing protocol²⁶ using the X-PLOR program²⁷. We used nuclear Overhauser effect (NOE)-derived distance restraints and hydrogen-bonding restraints, with a square-well potential and a force constant of $50.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The torsion restraints were used with a harmonic potential with a force constant of $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$.

Received 19 December 1997; accepted 17 March 1998.

- Chinnaiyan, A. M., O'Rourke, K., Tewari, M. & Dixit, V. M. FADD, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. *Cell* **81**, 505–512 (1995).
- Boldin, M. P. *et al.* A novel protein that interacts with the death domain of Fas/APO1 contains a sequence motif related to the death domain. *J. Biol. Chem.* **270**, 7795–7798 (1995).
- Alnemri, E. S. *et al.* Human ICE/CED-3 protease nomenclature. *Cell* **87**, 171 (1996).
- Muzio, M. *et al.* FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell* **85**, 817–827 (1996).
- Boldin, M. P., Goncharov, T. M., Goltsev, Y. W. & Wallach, D. Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell* **85**, 803–815 (1996).
- Hofmann, K., Bucher, P. & Tschopp, J. The CARD domain: a new apoptotic signalling motif. *Trends Biochem. Sci.* **22**, 155–156 (1997).
- Duan, H. & Dixit, V. M. RAIDD is a new 'death' adaptor molecule. *Nature* **385**, 86–89 (1997).
- Bertin, J. *et al.* Death effector domain-containing herpesvirus and poxvirus proteins inhibit both Fas- and TNFR1-induced apoptosis. *Proc. Natl Acad. Sci. USA* **94**, 1172–1176 (1997).
- Thome, M. *et al.* Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. *Nature* **386**, 517–521 (1997).
- Hu, S., Vincenz, C., Buller, M. & Dixit, V. M. A novel family of viral death effector domain-containing molecules that inhibit both CD-95- and tumor necrosis factor receptor-1-induced apoptosis. *J. Biol. Chem.* **272**, 9621–9624 (1997).
- Irmeler, M. *et al.* Inhibition of death receptor signals by cellular FLIP. *Nature* **388**, 190–195 (1997).
- Shu, H. B., Halpin, D. R. & Goeddel, D. V. Casper is a FADD- and caspase-related inducer of apoptosis. *Immunity* **6**, 751–763 (1997).
- Srinivasula, S. M. *et al.* FLAME-1, a novel FADD-like anti-apoptotic molecule that regulates Fas/TNFR1-induced apoptosis. *J. Biol. Chem.* **272**, 18542–18545 (1997).
- Goltsev, Y. V. *et al.* CASH, a novel caspase homologue with death effector domains. *J. Biol. Chem.* **272**, 19641–19644 (1997).
- Huang, B., Eberstadt, M., Olejniczak, E. T., Meadows, R. P. & Fesik, S. W. NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain. *Nature* **384**, 638–641 (1996).
- Liepinsh, E., Ilag, L. L., Otting, G. & Ibanez, C. F. NMR structure of the death domain of the p75 neurotrophin receptor. *EMBO J.* **16**, 4999–5005 (1997).
- Eberstadt, M., Huang, B., Olejniczak, E. T. & Fesik, S. W. The lymphoproliferation mutation in Fas locally unfolds the Fas death domain. *Nature Struct. Biol.* **4**, 983–986 (1997).
- Watanabe-Fukunaga, R., Brannan, C. I., Copeland, N. G., Jenkins, N. A. & Nagata, S. Lymphoproliferation disorder in mice explained by defects in Fas antigen that mediates apoptosis. *Nature* **356**, 314–317 (1992).
- Tartaglia, L. A., Ayres, T. M., Wong, G. H. & Goeddel, D. V. A novel domain within the 55 kd TNF receptor signals cell death. *Cell* **74**, 845–853 (1993).
- Keane, M. M., Ettenberg, S. A., Lowrey, G. A., Russell, E. K. & Lipkowitz, S. Fas expression and function in normal and malignant breast cell lines. *Cancer Res.* **56**, 4791–4798 (1996).
- Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N , ^{13}C , ^2H labeled proteins with high sensitivity. *J. Am. Chem. Soc.* **116**, 11655–11666 (1994).
- Clore, G. M. & Gronenborn, A. M. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Meth. Enzymol.* **239**, 349–363 (1994).
- Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J. Biomol. NMR* **3**, 225–231 (1993).
- Neri, D., Szyperki, T., Otting, G., Senn, H. & Wüthrich, K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry* **28**, 7510–7516 (1989).
- Vuister, G. W. & Bax, A. Quantitative ^1J correlation: a new approach for measuring homonuclear three-bond $^1\text{J}(\text{H}^{\text{N}}\text{H}^{\alpha})$ coupling constants in ^{15}N -enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777 (1993).
- Kuszewski, J., Nilges, M. & Brünger, A. T. Sampling and efficiency of metric matrix distance geometry: a novel partial metrization algorithm. *J. Biomol. NMR* **2**, 33–56 (1992).
- Brünger, A. T. *X-PLOR 3.2 Manual* (Yale Univ. Press, New Haven, 1992).
- Carson, M. J. Ribbon models of macromolecules. *J. Mol. Graph.* **5**, 103–106 (1987).

Acknowledgements. The 750-MHz NMR data were acquired at the National Magnetic Resonance Facility at Madison, which is supported by NIH, NSF and the University of Wisconsin.

Correspondence and requests for materials should be addressed to S.W.F. (e-mail: Stephen.Fesik@abbott.com). Coordinates for the NMR structures of the F25Y and F25G FADD death-effector domains have been deposited in the Brookhaven Protein Data Bank under the accession codes 1A1Z and 1A1W, respectively.

Crystal structure of the tetramerization domain of the *Shaker* potassium channel

Andreas Kreusch, Paul J. Pfaffinger*, Charles F. Stevens† & Senyon Choe

Structural Biology and † Molecular Neurobiology Laboratories, The Salk Institute, La Jolla, California 92037, USA

* Baylor College of Medicine, Houston, Texas 77030, USA

Voltage-dependent, ion-selective channels such as Na^+ , Ca^{2+} and K^+ channel proteins function as tetrameric assemblies of identical or similar subunits^{1–4}. The clustering of four subunits is thought to create an aqueous pore^{5,6} centred at the four-fold symmetry axis. The highly conserved, amino-terminal cytoplasmic domain (~130 amino acids) immediately preceding the first putative transmembrane helix S1 is designated T1. It is known to confer specificity for tetramer formation^{7,8}, so the heteromeric assembly of K^+ -channel subunits is an important mechanism for the observed channel diversity^{9–11}. We have determined the crystal structure of the T1 domain of a *Shaker* potassium channel at 1.55 Å resolution. The structure reveals that four identical subunits are arranged in a four-fold symmetry surrounding a centrally located pore about 20 Å in length. Subfamily-specific

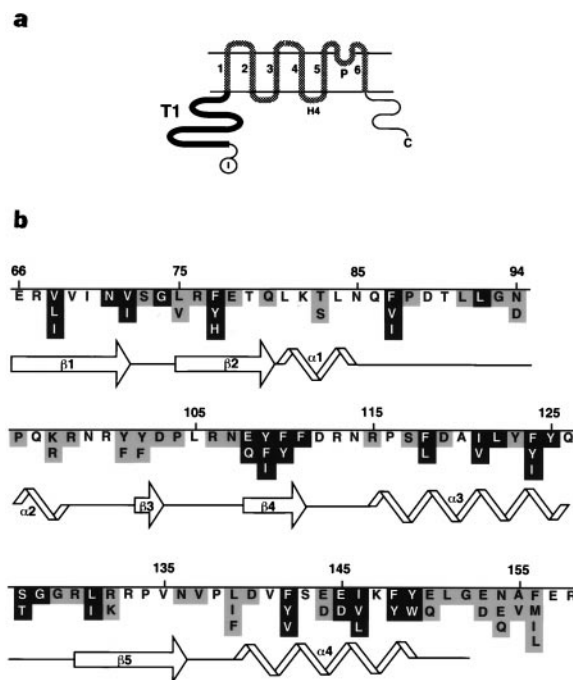


Figure 1 T1 monomer. **a**, Topology of the T1 segment with respect to the putative transmembrane helices. AKv1.1a is 515 amino acids long. The most conserved regions of the sequence are highlighted as thick lines: the T1 domain (amino acids 66–196) in black, and the transmembrane domain (197–440) in grey. S1–S6 are denoted by numbers, the P loop by P, and the inactivation ball (~30 amino acids) at the N terminus by I. **b**, Amino-acid sequence of AKv1.1 T1. Black boxes indicate residues that are conserved among 91 primary sequences from all four subfamilies (*Shaker*, *Shab*, *Shaw* and *Shal*); grey boxes, residues conserved only among *Shaker* subfamilies. For each conserved site, the substituting amino acids are listed. The unshaded amino acids are variable. The secondary structure assignment is given below the sequence (arrows for β -sheets; spirals for α -helices; solid lines for loops).

assembly is provided primarily by polar interactions encoded in a conserved set of amino acids at its tetramerization interface. Most highly conserved amino acids in the T1 domain of all known potassium channels are found in the core of the protein, indicating a common structural framework for the tetramer assembly.

The structure of the T1 monomer (amino acids 66–152; Fig. 1) consists of three distinct layers (Fig. 2a). The amino-terminal layer 1 (66–111) is formed by two pairs of antiparallel β -strands (β 1– β 2 and β 3– β 4) interrupted by two short α -helices (α 1– α 2) between them. The β 4 is hydrogen-bonded in parallel to the carboxy-terminal half of β 1, forming a four-stranded β -sheet of layer 1. The loop between β 1 and β 2 forms a type II glycine turn. Following layer 1 is a 12-residue-long α -helix (α 3) constituting layer 2 (115–126). A single β -strand (β 5) and an α -helix (α 4) form C-terminal layer 3 (130–152). A structural homology search using the DALI¹² and SCOP¹³ databases reveals that the fold of the T1 domain is different from any previously determined fold.

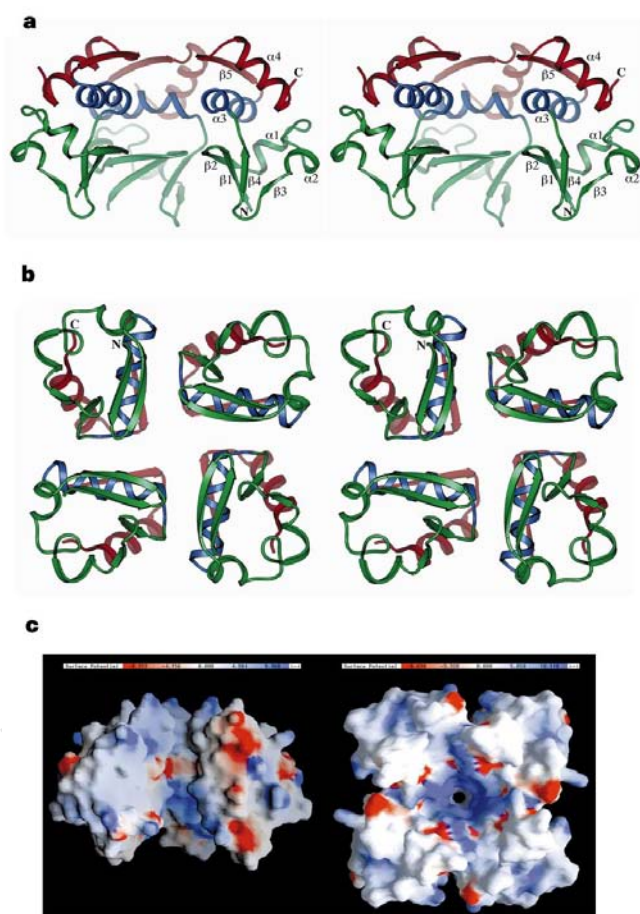


Figure 2 T1 tetramer **a**, Stereo side-view of the T1 tetramer. Only three subunits of the tetramer are shown; the frontmost subunit is omitted for clarity. The four-fold symmetry axis is vertical. Layers 1, 2 and 3 in each subunit are shown in green, blue and red, respectively, starting from the N terminus at the putative cytoplasmic side. The α -helical and β -sheet segments are labelled as in Fig. 1b. All figures except Figs 1 and 2c have been prepared using SETOR²⁹. **b**, Stereo view of the T1 tetramer from the N-terminal side along the four-fold axis, which passes through the centre of the tetramer. **c**, Molecular surface representation of the T1 tetramer. Electrostatic potential calculated by GRASP³⁰ is colour coded on the surface from blue ($\sim$ 10 kT) to red ($\sim$ 7 kT). Left, side view as in **a**. Because of the symmetry, the surface visible on the subunit to the left of the pore interfaces with a surface on the omitted subunit, which is identical to the subunit surface visible on the right. Right, the T1 tetramer is viewed from the cytoplasmic side as in **b**. The narrowest opening ($\sim$ 3.2 Å in diameter) in the centre of the pore (white region) is formed by side-chain atoms of N136.

Four subunits assemble around a four-fold axis (Fig. 2b). The three layers are stacked sequentially along the direction of the pore axis. The N and C termini of each subunit are located at opposite faces of the tetramer, suggesting that the N-terminal layer 1 faces the cytoplasm and the C-terminal layer 3 faces the membrane. This interpretation is based on the proximity between the C terminus of T1 and the N terminus of S1, as the path of the 38 C-terminal amino acids (Arg 160–Leu 197) removed by trypsinization cannot be precisely predicted. In the case of the T1 structure of the *Shaw* channel, a further 15 C-terminal amino acids continue to fold on the membrane-facing side of T1, forming layer 4 (K. Bixby *et al.*, manuscript in preparation). At the cytoplasmic side of the tetramer, four protruding loops between β 3 and β 4 form prominent extensions and create a cradle-shaped depression 15 Å deep, whereas the membrane-facing side of the tetramer is relatively flat (Fig. 2c).

An alignment of available T1 domain sequences for voltage-gated potassium channels (54 K_v1, 17 K_v2, 11 K_v3 and 9 K_v4 sequences) shows that 24 positions are highly conserved including seven absolute identities (Asn 71, Gly 74, Leu 92, Phe 111, Leu 122, Tyr 125 and Gly 128; see Figs 1b and 3a). Most of these highly conserved residues are located in the hydrophobic core of the T1 monomer, suggesting that the overall fold of T1 is well preserved among potassium channel proteins. Several K⁺ channels have insertions ranging in length from four (mouse K_v3.1) to 51 residues (K_v3.2 sequences) between the conserved subdomains A and B. These insertions occur at the C terminus of α 2 (Fig. 3a), and so are not likely to participate in subunit interactions.

Within the T1 tetramer, each subunit buries 940 Å² or 20% of its solvent-accessible surface area at its two subunit interfaces. Inter-

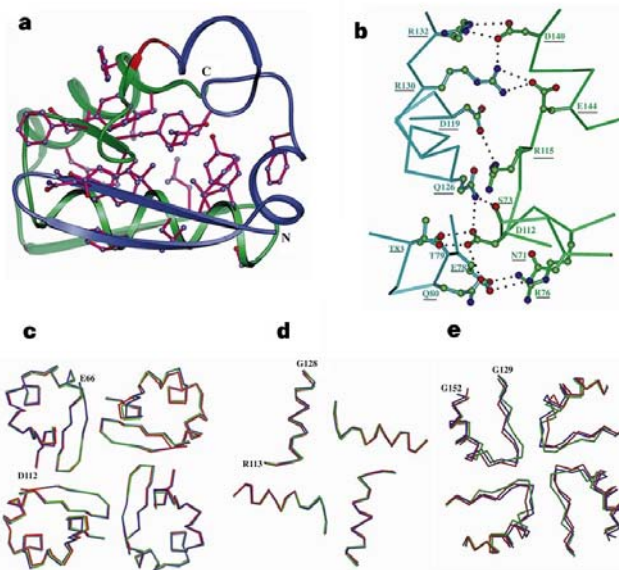


Figure 3 T1 side chains and views of layers 1, 2 and 3. **a**, Side chains conserved in all four subfamilies of voltage-gated potassium channels (those in black boxes in Fig. 1b). The backbone of the T1 subunit is shown as a ribbon. Subdomain A is in blue, the variable region between subdomains A and B is in red, and subdomain B is in green. **b**, Side chains of 15 residues involved in polar intersubunit interactions (E 78, T 79, Q 80, T 83, D 119, Q 126, R 130 and R 132 from subunit 1; N 71', S 73', R 76', D 112', R 115', D 140' and E 144' from subunit 2). Only side chain-side chain interactions are shown. Underlined residues are conserved in all *Shaker* subfamily members. **c–e**, Cross-sectional views of layers 1 (**c**), 2 (**d**) and 3 (**e**). C α backbone atoms of the T1 tetramer in the three crystal forms (I in blue, C in green and P in red) are superimposed. Based on all main-chain atoms, the root mean square deviations are 0.92 Å between the P and C forms, 0.84 Å between the I and C forms, and 0.77 Å between the I and P forms. The most significant differences occur within the region between R 133 and D 140 (r.m.s. deviations between 1.25 and 2.1 Å).

actions between subunits are predominantly polar, including five salt bridges (Arg 76–Glu 78, Arg 115–Asp 119, Arg 130–Asp 140, Arg 130–Glu 144 and Arg 132–Asp 140), one main chain/side chain (Arg 76–Arg 115), one main chain/main chain (Ser 73–Glu 78) and six side chain/side chain hydrogen bonds (Asn 71–Glu 78, Ser 73–Gln 126, Thr 79–Asp 112, Gln 80–Asp 112, Thr 83–Asp 112 and Asp 112–Gln 126), as well as several water-mediated hydrogen bonds. These polar interactions involve 15 different side chains, which are uniformly distributed over the interface area (Fig. 3b).

One of the contacts within layer 1 of the tetramer is mediated by a hydrogen bond between main-chain atoms (the carbonyl O of Ser 73 within the glycine turn of $\beta 1$ – $\beta 2$ loop and the amide N of Glu 78 of $\beta 2$ of a neighbouring subunit). Gly 74, which is 100% conserved among all T1 sequences, provides an essential structural role for the formation of this hydrogen bond (Fig. 3c). In layer 2, four $\alpha 3$ helices are mutually perpendicular to each other with the N-terminal end (Arg 115) packing against the side of a neighbouring helix (Asp 119) (Fig. 3d). The relative orientation of the helical dipoles¹⁴ will be perpendicular to the direction of the local electric field of the membrane. Interactions between layer 3 are highly polar, mediated by three sets of charge pairs (Arg 130–Asp 140, Arg 130–Glu 144 and Arg 132–Asp 140). Layers 1 and 2 are almost identical among all three crystal forms. However, layer 3 loops show considerable differences in their backbone conformation (Fig. 3e), which do not result from crystal packing, as this loop region is not involved in direct packing contacts in any crystal form.

There is a centrally located, water-filled pore along the four-fold axis of the T1 tetramer (Fig. 2c). Its physical shape and the location of T1, adjacent to the cytoplasmic inner leaflet of the membrane, suggest that it is an excellent candidate to form the cytoplasmic vestibule of the ion-conducting pathway of the potassium channel. The narrowest opening of the pore is formed by C β of Asn136 within the flexible loop of layer 3 at the membrane side of the T1 tetramer. Within this pore, at least 20 water molecules are common between the three different crystal forms. They are arranged in a four-fold rotation symmetry with separations ranging from 4.8 to 8.4 Å (data not shown). If this putative pore forms an entrance to

the transmembrane region of the pore, the four-fold symmetry of these water molecules may be to substitute for four hydration waters of the central plane of the normally octahedrally coordinated ion traversing the pore. Whether the observed pore in the centre of the T1 tetramer is part of the ion-conducting pathway is open to experimental verification.

Side chains involved in polar interactions at the subunit interface are conserved in a subfamily-specific manner (Fig. 3b). For 15 positions involved in polar intersubunit contacts, 54 *Shaker*-like sequences show that 13 positions are either absolutely conserved (Asn 71, Ser 73, Arg 76, Glu 78, Gln 80, Arg 115, Asp 119, Arg 130 and Asp 140) or conservatively changed (Thr 83 to Ser, Gln 126 to His, Arg 132 to Lys and Glu 144 to Asp). The remaining two are each replaced only once non-conservatively (Thr 79 to Ala in dog K_v1.5 and Asp 112 to Val in dog K_v1.2). This level of conservation is significantly higher than that calculated for the T1 sequences of K_v1 channel proteins (~68%). Most of these positions (8 out of 15) are absolutely conserved only among *Shaker* sequences. From the total of 37 non-*Shaker* sequences (K_v2, K_v3 and K_v4), we find only one residue (Asn 71) common with *Shaker* channels. This occurrence of 6.7% is in contrast to the 39% overall chance of finding positions conserved in both non-*Shaker* and *Shaker* channels (24 out of 62 *Shaker*-conserved positions; see Fig. 1b). At these positions among non-*Shaker* channels, only four are partly conserved (Gln 126 to Arg or Lys, Arg 130 to Lys or Gln, Arg 132 to His, and Glu 144 to Asp or Gln), and the remaining nine are highly variable. Therefore, such a distinctive pattern of sequence conservation at selected positions within a subfamily strongly suggests that a subset of these 15 polar amino acids can encode *Shaker*-specific determinants for homomeric as well as for heteromeric assembly predominantly through polar interactions.

T1 will interact with other parts of the channel molecule. A single molecule of an inactivation peptide^{15,16} originating from the N terminus of a channel subunit (N-type inactivation) binds to a structurally conserved^{17,18} region of the channel. A putative binding site for this peptide is the β -wheel of layer 1 surrounded by four protruding loops which form a cradle-like structure, 25 Å in diameter and 15 Å (C α distances) in height (Fig. 2c). The size and

Table 1 X-ray data and phasing/refinement statistics

Data set*	Crystal form	Space group	Resolution (Å)/ redundancy	Completeness (%) (All/outer shell)	R_{sym} (%)† (All/outer shell)
NATIVE	I	I4	1.80/10.9	98.7/92.3	6.3/45.0
R132C	C	C222 ₁	1.55/10.0	98.7/99.0	4.0/10.3
N154C-1	P	P2 ₁ 2 ₁ 2 ₁	1.95/6.7	99.4/99.5	6.0/14.7
N154C-2	P	P2 ₁ 2 ₁ 2 ₁	2.20/15.3	97.1/98.9	2.5/9.6
MAD phasing statistics for N154C-2					
Resolution (Å)	50–10.0	10.0–5.63	5.63–3.92	3.92–3.00	Overall
Phasing power‡					
$\lambda_1 = 1.0507$ Å					
Anomalous	1.45	0.88	0.53	0.34	0.54
Isomorphous	–	–	–	–	–
$\lambda_2 = 1.0059$ Å					
Anomalous	1.79	1.33	0.89	0.62	0.88
Isomorphous	1.23	1.35	0.86	0.66	0.90
$\lambda_3 = 1.0094$ Å					
Anomalous	1.80	1.20	0.73	0.49	0.74
Isomorphous	1.31	1.09	0.67	0.45	0.63
Mean figure of merit§	0.69	0.62	0.44	0.33	0.44
Refinement statistics					
Crystal form	Resolution (Å)	No. of protein atoms (water)	Average B-factors (Å ²) (protein/water)	$R_{\text{factor}}/R_{\text{free}}$ (%)¶	
I	10–1.8	742 (35)	25/40	20.0/24.2	
C	26–1.55	1,482 (128)	14/24	20.4/24.4	
P	50–2.0	2,984 (164)	19/35	26.3/34.5	

* Data sets NATIVE and N154C-2 were collected at SSRL 1–5; R132C and N154C-1 at SSRL 9–1.

† $R_{\text{sym}} = 100 \times \sum_i |I_i - \langle I_i \rangle| / \sum_i I_i$.

‡ Phasing power is r.m.s. of the heavy-atom structure factor amplitude divided by lack of closure. Phasing statistics are based on two mercury sites per asymmetric unit.

§ Figure of merit = $(|\sum P(\alpha)| / \sum P(\alpha))$, where $P(\alpha)$ is the phase probability distribution and α is the phase.

|| Number of protein atoms and water molecules per asymmetric unit (1 for I; 2 for C, and 4 for P).

¶ R_{free} was calculated with 5% (NATIVE, N154C-1), and 7.5% (R132C) of the data. No sigma cutoff was applied.

overall shape of this cradle seems to be appropriate to accommodate a single Raw3 inactivation peptide¹⁹, which would effectively block the T1 pore entrance. Although this view is structurally attractive, experimental data suggest a role for the 14-residue-long H4 loop, linking S4 and S5, in binding the inactivation ball²⁰. Thus it is conceivable that the ball may access the loop near the outer rim of the membrane-facing side of T1 tetramer to confer its action.

The flexibility of the Pro 134–Pro 138 loop (Fig. 3e) on the membrane-facing side of the T1 tetramer might result from the absence of neighbouring atoms such as subdomain C, H4 loop linking S4 and S5, and the C-terminal tail of S6. Gating-dependent access of Cd²⁺ to the C terminus of S6 indicates that the gating elements reside between the C terminus of S6 and the cytoplasmic part of the channel²¹. During channel activation, S4 shifts significantly (~10 Å) in an outward direction in response to a depolarizing membrane potential²². Such a large conformational change in S4 can possibly trigger a series of structural changes in the membrane-facing region of T1 tetramer, which may involve the flexible loop of layer 3.

Our structural studies of the T1 domain demonstrate the four-fold symmetric arrangement of potassium-channel subunits; a long, narrow pore structure forming the putative inner vestibule of the ion-conducting pathway; a highly conserved fold of the T1 domain; and the structural basis for subfamily-specific assembly of potassium-channel proteins. Our results will provide a basis for further systematic analyses by site-specific mutagenesis. □

Methods

Protein purification and expression. The wild-type T1 domain and two single-point mutant proteins (R132C and N154C), each comprising residues 65–197 (S1 being Leu 197–Phe 221) including an N-terminal Met (residue 65) (AK₁L1), were expressed as fusion proteins (pET 20b from Novagen) in *Escherichia coli* BL21. This T1 segment contains three highly conserved subdomains: A (residues Arg 67–Gln 96), B (Tyr 102–Lys 171) and C (Gln 179–Arg 196)⁷. The proteins were purified by Ni-NTA affinity chromatography (Qiagen) in 6 M urea, and refolded by stepwise removal of urea. After refolding, trypsinization removed 38 C-terminal residues (subdomain C). The truncated T1 protein contained residues 65–159 (Fig. 1b), as determined by mass spectroscopy and N-terminal sequencing. Results from gel filtration (Superdex 75, Pharmacia), dynamic light scattering (Biotage) and ultracentrifugation (Beckman) showed that the purified T1 protein forms a stable homotetramer.

Crystallization and data collection. (1) I crystal form. Crystals of the wild-type T1 domain have the space group *I*4 (*a* = *b* = 51.03 Å, *c* = 65.61 Å) with one monomer per asymmetric unit. They were grown by vapour diffusion at 15 °C with a precipitating solution containing 7–8% PEG8000, 300 mM NaCl, 200 mM MgCl₂, 20 mM Tris (pH 8.5). (2) C crystal form. Crystals of R132C were obtained in the space group *C*222₁ (*a* = 65.64 Å, *b* = 68.47 Å, *c* = 88.40 Å) with two monomers per asymmetric unit from drops containing 3.5% PEG8000, 100 mM TAPS (pH 8.5), 250 mM LiSO₄, 300 mM NaCl. (3) P crystal form. N154C crystallized in space group *P*₂₁₂₁2₁ (*a* = 56.98 Å, *b* = 69.43 Å, *c* = 91.54 Å) with four monomers per asymmetric unit, after the protein was preincubated with 10–50 mM ethyl mercury chloride for 4 h. X-ray diffraction data were collected at the Stanford Synchrotron Radiation Laboratory or with an in-house imaging-plate detector (DIP2020k) running on a rotating anode. The data set NATIVE was collected at room temperature. Data sets R132C, N154C-1 and N154C-2 were collected from flash-frozen crystals at 100 K. All data were processed with DENZO and SCALEPACK²³.

Phase determination and refinement. Phase information was derived from multiwavelength anomalous diffraction (MAD)²⁴ data collected at three wavelengths using the inverse beam technique from a mercury derivative of a N154C mutant (N154C-2 data; see Table 1). Initial phases calculated with MLPHARE²⁵ to 3.0 Å resolution (Table 1) were improved by iterative solvent flattening, non-crystallographic symmetry averaging and skeletonization, using DM²⁵. The resulting electron-density maps were of sufficient quality to build a model including most of the side chains, using the program O²⁶. After one round of simulated annealing refinement (XPLOR 3.8)²⁷ using non-

crystallographic symmetry restraints, this partly refined model was used to solve forms I (NATIVE) and C (R132C) by molecular replacement methods. Both XPLOR 3.8 and Amore²⁸ programs gave unambiguous results. All models in the three crystal forms were independently refined by simulated annealing, positional refinement and isotropic *B*-factor refinement using XPLOR 3.8. After the resolution for refinement was increased to 2.2 Å, the non-crystallographic symmetry restraints for the protein models for the C and P forms were released. The final structural models contain residues 66–152. Residues at neither termini are visible, probably as a result of their high mobility. Water molecules were independently added in all three forms (I, C and P). In all three models, the main-chain conformations of all residues lie in favoured and allowed regions of ϕ/ψ in a Ramachandran plot.

Received 29 January; accepted 9 March 1998.

- Noda, M. *et al.* Primary structure of *Electrophorus electricus* sodium channel deduced from cDNA sequence. *Nature* **312**, 121–127 (1984).
- Tanabe, T. *et al.* Primary structure of the receptor for calcium channel blockers from skeletal muscle. *Nature* **328**, 313–318 (1987).
- Papazian, D. M., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. Cloning of genomic and complementary DNA from Shaker, a putative potassium channel gene from *Drosophila*. *Science* **237**, 749–753 (1987).
- MacKinnon, R. Determination of the subunit stoichiometry of a voltage-activated potassium channel. *Nature* **350**, 232–235 (1991).
- Li, M., Unwin, N., Stauffer, K. A., Jan, L. Y. & Jan, Y. N. Images of purified *Shaker* potassium channels. *Curr. Biol.* **4**, 110–115 (1994).
- Miller, C. 1990: Annus mirabilis of potassium channels. *Science* **252**, 1092–1095 (1991).
- Shen, N. V., Chen, X., Boyer, M. M. & Pfaffinger, P. J. Deletion analysis of K⁺ channel assembly. *Neuron* **11**, 67–76 (1993).
- Li, M., Jan, Y. N. & Jan, L. Y. Specification of subunit assembly by the hydrophilic amino-terminal domain of the *Shaker* potassium channel. *Science* **257**, 1225–1230 (1992).
- Schwarz, T. L., Tempel, B. L., Papazian, D. M., Jan, Y. N. & Jan, L. Y. Multiple potassium-channel components are produced by alternative splicing at the *Shaker* locus in *Drosophila*. *Nature* **331**, 137–142 (1988).
- Wei, A. *et al.* K⁺ current diversity is produced by an extended gene family conserved in *Drosophila* and mouse. *Science* **248**, 599–603 (1990).
- Stühmer, W. *et al.* Molecular basis of functional diversity of voltage-gated potassium channels in mammalian brain. *EMBO J.* **8**, 3235–3244 (1989).
- Holm, L. & Sander, C. Protein structure comparison by alignments of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- Murzin, A. G., Brenner, S. E., Hubbard, S. T. & Chothia, C. "SCOP": a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540 (1995).
- Hol, W. G. van Duijnen, P. T. & Berendsen, H. J. The alpha-helix dipole and the properties of proteins. *Nature* **273**, 443–446 (1978).
- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of *Shaker* potassium channel inactivation. *Science* **250**, 533–538 (1990).
- MacKinnon, R., Aldrich, R. W. & Lee, A. W. Functional stoichiometry of *Shaker* potassium channel inactivation. *Science* **262**, 757–759 (1993).
- Foster, C. D., Chung, S., Zagotta, W. N., Aldrich, R. W. & Levitan, I. B. A peptide derived from the *Shaker* K⁺ channel produces short and long blocks of reconstituted Ca²⁺-dependent K⁺ channels. *Neuron* **9**, 229–236 (1992).
- Toro, L., Stefani, E. & Latorre, R. Internal blockade of a Ca²⁺-activated K⁺ channel by *Shaker*B inactivating "ball" peptide. *Neuron* **9**, 237–245 (1992).
- Antz, C. *et al.* NMR structure of inactivation gates from mammalian voltage-dependent potassium channels. *Nature* **385**, 272–274 (1997).
- Isacoff, E. Y., Jan, Y. N. & Jan, L. Y. Putative receptor for the cytoplasmic inactivation gate in the *Shaker* K⁺ channel. *Nature* **353**, 86–90 (1991).
- Liu, T., Holmgren, M., Jurman, M. E. & Yellen, G. Gated access to the pore of a voltage-dependent K⁺ channel. *Neuron* **19**, 175–184 (1997).
- Mannuzzo, L. M., Moronne, M. M. & Isacoff, E. Y. Direct physical measure of conformational rearrangement underlying potassium channel gating. *Science* **271**, 213–216 (1996).
- Otwiński, Z. in *Data Collection and Processing* (eds Swayner, L., Isaac, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, Warrington, UK, 1993).
- Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
- Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Crystallography. *Acta Crystallogr. D* **50**, 670–673 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. W. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. X-PLOR Version 3.8 (Dept of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 1996).
- Navaza, J. Amore: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Evans, S. V. SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134–138 (1993).
- Nicholls, A., Sharp, K. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Acknowledgements. We thank H. Bellamy at beamline 1-5 at SSRL for his help; W. Yang, J. Greenwald and R. Robinson for data collection; P. Voelker for analytical centrifuging; A. Craig for mass spectroscopy; W. Fischer and C. Park for amino-acid sequencing; and G. Louie, A. Bilwes, K. Bixby, J. Noel, F. Crick and L. Orgel for discussions and comments on the manuscript. A.K. acknowledges fellowships from the Hoffman Foundation and the American Heart Association. SSRL is funded by the Department of Energy, Office of Basic Energy Science. This work was supported by grants from the NIH (P.J.P. and S.C.) and the Council for Tobacco Research (S.C.). S.C. is a Neuroscience fellow of the E.A. & J. Klingenstein Fund. C.E.S. is a HHMI investigator.

Correspondence and requests for materials should be addressed to S.C. The atomic coordinates have been deposited in the Brookhaven Databank, accession code 1a68.

DNA discourse

Molecular biology is the subject and the objects of attention include a gel electrophoresis comet assay, an adenovirus expression kit, proof-reading polymerases, and a large-scale mouse cDNA expression array.

SFM 25

Kontron Instruments

A spectrofluorometry instrument with WIND 25 software

Using Microsoft Windows as its platform, WIND 25 is a 32-bit application that allows fast calculations and multitasking operations. It is designed for user friendliness, with a graphical interface that guides the user through all the functions and features of the system. Advanced calculations offered with this system include enzyme determination and substrate, Michaelis-Menten, ratio and purity calculations. The instrument uses a high-intensity xenon light source, a double-beam optical system with blazed gratings and a high-sensitivity photomultiplier, to provide higher sensitivity and consistent, reproducible results.

Reader Enquiry No. 100

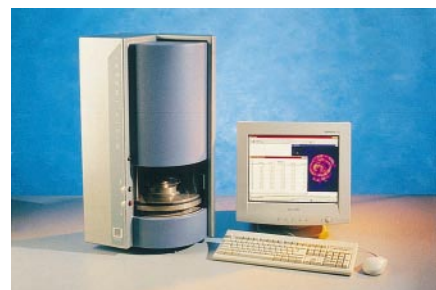
Micro Imager

From Bio Image

Fast tritium detection in ISH experiments

Developed by Biospace Mesures in France, this instrument detects tritium or other β -emitting isotopes in tissue sections and *in situ* hybridization (ISH) experiments up to $\times 50$ faster than film, says Bio Image. The system achieves a spatial resolution of 15 μm and is also capable of dual-isotope detection. It can simultaneously detect tritium and another isotope in the same sample, thus eliminating the error when overlaying two samples. The system is used for β detection in applications, such as ISH experiments, immunocytochemistry, receptor-ligand binding assays and metabolic tracer experiments. The Micro Imager detects all β -emitting isotopes with a detection efficiency of 60–90 per cent and background noise less than 0.01 c.p.m. mm^{-2} . Real-time imaging allows the user to stop the acquisition at any time, reducing the risk of over-exposure or under-exposure.

Reader Enquiry No. 101



Bio Image's fast tritium detection system.

CometSlide

From Trevigen

A single-cell gel electrophoresis (comet) assay

The assay kit allows the quantitative evaluation of DNA fragmentation associated with DNA damage or apoptosis. It provides low melting point (LMP) agarose, lysing buffer, nucleic acid stain and the company's CometSlide — a specially treated microscope slide, which allows direct application of cells in LMP agarose to the surface, without pre-treatment. The CometSlide features a specially designed hydrophobic barrier for incubating cell samples with Trevigen's DNA repair enzymes, allowing analysis of specific types of gene damage.

Reader Enquiry No. 102

GS2000

From UVP

A DNA fragment analyser for automated fluorescence, sizing, sequencing and quantification

This system utilizes an ultra-thin gel apparatus and is stated to have the resolution of single-channel capillary systems, but with the advantage of being able to run many samples in parallel. Gels are poured *in situ* and can be re-used up to five times. Samples from 50–100 μl can be loaded with standard pipettors using 48- and 64-well combs. Real-time electrophoresis detection will analyse fluorescently labelled DNA as it migrates through the gel. The sensitivity and resolution of the system is said to allow the system to detect and size 100 attomoles of DNA in a single peak. The system is said to cost less to operate, owing to reagents savings. It will detect most commercially available fluorophores. For non-denaturing applications ethidium bromide can be used to detect as little as 500 attomoles of DNA.

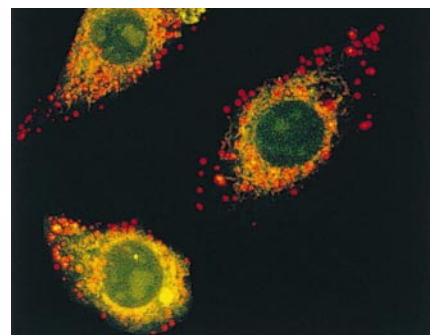
Reader Enquiry No. 103

Mirus' Label It

From Pan Vera

Kits for efficient one-step, non-radioactive labelling of DNA or RNA with a single reagent

This method is said to tag covalently and non-destructively original substrate nucleic acids. In contrast to some laborious and poorly controlled enzymatic reactions, these reactions are said to be simple, non-enzymatic and result in reproducible products and yields. In addition, the reactions can be scaled in a single tube without loss of efficiency. Products can be used in many molecular biology applications, including FISH and



Label It: non-radioactive labelling from Pan Vera.

DNA or RNA transport studies. The kits are available for biotin, digoxin, rhodamine and fluorescein labelling applications.

Reader Enquiry No. 104

Adenovirus expression vector kit

From TaKaRa

A more efficient means of generating recombinant adenovirus for gene transfer that is based on the COS-TPC method

Adenovirus vectors have not been widely used for gene transfer because the procedure for generating adenovirus vectors is laborious and inefficient. The COS-TPC method utilizes genomic DNA-terminal protein complex (DNA-TPC) and a cosmid with the target gene inserted. Whereas, conventional methodology uses a plasmid and adenovirus genomic DNA without terminal protein (TP), though both terminals of the adenovirus genomic DNA are originally tagged with TP. According to the manufacturer, high-titre virus can be obtained using the COS-TPC method (10^8 – 10^9 virus per millilitre), which can be further concentrated to 10^{11} pfu ml^{-1} . This is said to result in a system that can transfer the target gene to nearly 100 per cent of adherent cells. It is available for a wide range of animals and for infecting both resting cells and proliferating cells. It can be targeted to a number of differentiated and undifferentiated cultured animal cells, including nervous system cells.

Reader Enquiry No. 105

Image

From Life Science Resources

An image analysis and processing platform available for gene imaging and other molecular biology applications

This platform uses the latest fluorescence probes, such as green fluorescent protein (GFP) and chemiluminescent probes (aequorin and luciferase), to provide an integrated system for the study of gene

expression in living and dead cells. Imagen can acquire either single or time-lapse images labelled with one or more fluorescent probes, specific to the gene or genes of interest. Image analysis tools provide image montages, intensity plots with time graphical overlays that relate fluorescence to gene activity. The system is fully integrated with the company's 16-bit digital cameras available with resolution up to $1,317 \times 1,035$ pixels. Back-thinned CCDs are available for studies utilizing the low-light emitting chemiluminescent probes, aequorin and luciferase. Automated control of external hardware is provided as an integral part of the system, including a high-speed filter wheel, a monochromator, z-stepper stage control and camera shutter. The image processing digital deconvolution software is also available as an optional upgrade to the system. This image enhancement capability can smooth, sharpen or remove out-of-focus haze from two- or three-dimensional microscope images.

Reader Enquiry No. 106

TGGE

From Biometra

Rapid mutation detection by temperature-gradient gel electrophoresis (TGGE)

According to the manufacturer, this new instrument is useful for both mutant screening and allele typing. It uses a specially designed electrophoresis unit that utilizes Peltier elements for precise control of the chamber temperature. In addition, a combined system controller and power supply provides the run conditions for gradient or standard temperature profiles, which is necessary for DNA migration. TGGE is said to be an improvement over other DNA screening methods in that it can detect almost 100 per cent of all mutations, says Biometra.

Reader Enquiry No. 107

GDS-20

From ATR

A gel drying system that provides efficient oil-free drying at an economical price

The system incorporates the HGD-10 large-slab gel dryer, the GDTT inlet/outlet trap system, a TD-20 diaphragm pump, and all the connections and tubing required for assembly. Solid, Teflon pump heads, continuous moulded diaphragms and a multiport valve design are said to provide greater reliability than other diaphragm pumps. Unlike water aspirators, diaphragm pumps maintain a constant vacuum, resulting in more consistent gel drying. The expansion flask and charcoal filter provide a vapour-free, odour-free environment and isolate radioactive waste. (Charcoal filters are inexpensive and easy to replace.) Designed for convenience, as well as performance, the



ATP offers its GDS-20 gel drying system.

system can be installed and ready for use in minutes, says ATR.

Reader Enquiry No. 108

GeneGenius

From Syngene

A complete gel documentation and analysis system for molecular biology

GeneGenius is a PC-based configuration with a darkroom, camera and transilluminator, as well as the GeneSnap acquisition system, GeneTools one-dimensional analysis software and a thermal printer. These modules provide image acquisition and processing for laboratories that are working with DNA and protein gels, autoradiographs and other types of gel media. The system combines a simple user interface and is supplied with a PC that has a CD-ROM drive, keyboard, mouse and a Windows95 or Windows NT user interface. Included with this package are word-processing, spreadsheet and publishing programs so that results and data can be exported for publication or analysis. GeneTools analysis software is also available as a stand-alone package for use with images captured using different systems. It handles a wide range of molecular biology applications: agarose gels, Southern blots, western blots, northern blots, autorads and SDS-PAGE. An ELISA plate/dot-blot and colony counting options are also available. GeneGenius is also said to perform band quantification, molecular size determination, band matching, band comparison, volume measurements, R_f determination, dendograms and profiling comparisons.

Reader Enquiry No. 109

96-well gel electrophoresis

From MadgeBio

Using Madge gels, this system is designed for molecular biology reactions in 96-well plates

Developed in the UK, this system utilizes existing gel electrophoresis tanks and is said to reduce labelling confusion and transfer times from the reaction plate to the gel. High throughput can be achieved in a shoe-box-sized gel tank, processing up to several thousand samples in a few hours, says MadgeBio. High resolution is achieved in the 20–1,000-bp range. Moreover, automation is now possible and high-quality analytical software is available to analyse the resulting gels. The gel-plate measures

170 mm $\times$ 110 mm and fits into conventional gel electrophoresis tanks with 96 8- μ l wells. The gel is set at an angle to the field, and provides a 2.6-cm separation run — adequate separation for most applications. The kit provides the means to produce either agarose or acrylamide gels.

Reader Enquiry No. 110

FTA Gene Guard

From Life Technologies

For collection, storage and purification of blood samples for genetic analysis

This technique allows simple collection and long-term, room-temperature storage of blood samples for human identity testing and genetic analysis. The FTA-treated matrix protects DNA during storage and allows simple purification of high-quality genomic DNA. Matrix-bound DNA enables consistent PCR and RFLP results, and is said to be well suited for high-throughput applications. The FTA card consists of filter paper impregnated with a proprietary formulation of denaturants, which are designed to maintain the integrity of the DNA contained in the specimen, and to prevent growth of bacteria and microorganisms. DNA can be isolated from blood (up to about 500 μ l), buccal swabs, epithelial cells, bacteria, viruses and cultured cells. Once biological fluid is spotted on the FTA card, DNA purification takes less than an hour, states the manufacturer. Blood spotted on the card does not require refrigeration, and can be stored indefinitely at room temperature in a typical filing system, eliminating the need for large freezers.

Reader Enquiry No. 111

Reagents

India-HRP

From Pierce

A new probe, together with its metal-activated derivatives for the detection of target molecules

This derivative of horseradish peroxidase allows metals to be bound in an active form for subsequent interaction and detection of target molecules. The active chelator has similar binding capabilities to those reported for iminodiacetic acid. The India HisProbe-HRP, a nickel-activated derivative of HRP, has been optimized for direct detection of recombinant histidine-tagged proteins and other histidine-rich proteins. India PhosphoProbe-HRP, an iron-activated derivative of HRP, is phosphate ($R-PO_3$)-specific when coupled with a novel treatment termed 'reactive chemical binding' (RCB). This probe has been optimized for direct detection of phosphoester ($R-PO_3$) molecules, such as nucleotides or protein/peptides containing phosphoserine, phosphothreonine and phosphotyrosine.

Reader Enquiry No. 112



Hyped on hypochlorite solution for cell analysers.

Hypo Chlorite

From Mallinckrodt Baker

An improved hypochlorite reagent that may mean longer life, simpler cleaning procedures and reduced down-time for cell analysers

By adding special tube-protecting compounds to its traditional product, the manufacturer may help to ensure users of continued malleability of plastic tubes and components together with suppressed noise. This new, odourless hypochlorite not only offers more effective cleaning, according to the manufacturer, but is considerably less corrosive than alternative products.

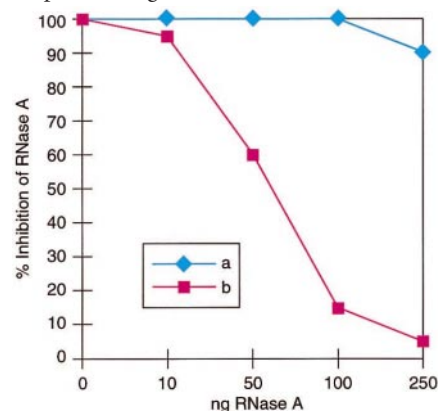
Reader Enquiry No. 113

Prime RNase inhibitor

From Flowgen

A human placental inhibitor that is said to reduce costs by 50 per cent

According to the manufacturer, this is a broad specificity, protein-based RNase inhibitor that is 50 times more active at inhibiting RNase A, B and C than most human placental inhibitors. It is said to exhibit activity and stability over a broad range of DTT concentrations, temperature and pH (5.0–9.0). It is stable at elevated temperatures (37 °C) for as long as two hours. Moreover, this inhibitor is suitable for protecting RNA samples from degradation in cDNA synthesis; *in vitro* transcription using the SP6, T7 and HeLa cell



a: 1 unit PRIME RNase INHIBITOR™
b: 30 units Placental RNase Inhibitor™

Potency of Prime RNase inhibitor from Flowgen.

extract systems; in *in vitro* translations that use rabbit reticulocyte lysates and wheat germ extracts. Additionally, Prime RNase inhibitor can prevent RNA degradation in cell extracts prepared under non-denaturing conditions. It can be utilized in first-strand synthesis before PCR and in the RT-PCR of RNA isolated from single cells. This reagent should also find use in gel retardation assays, the isolation of an RNA binding protein by oligonucleotide affinity purification and in nuclear run-on transcription assays.

Reader Enquiry No. 114

Complementary tools

Dynabeads Oligo (dT)₂₅

From Dynal

Beads for mRNA isolation and subsequent production of subtracted cDNA probes

This method can be used for the screening and isolation of rare, differentially expressed mRNAs. The first step is the isolation of polyA⁺ RNA, using Dynabeads Oligo (dT)₂₅. Then, instead of eluting the captured mRNA, a solid-phase cDNA library is constructed directly on the bead's surface, using the oligo(dT) sequence bound to the bead as a primer. This library is specific to a particular cell type or tissue. Next, subtracted first-strand cDNA is immobilized on beads and this is hybridized with mRNA from the target material, leaving subtracted mRNA in the supernatant after the removal of common mRNA captured with the Dynabeads. Unique mRNA is subsequently reverse transcribed to radiolabelled cDNAs and used to screen cDNA libraries. In addition, both first- and second-strand cDNA are synthesized for cDNA cloning. Magnetic handling is said to minimize losses at each step, and also enable simple and rapid buffer changes to optimize conditions for hybridization and specific enzymatic reactions. The subtracted cDNA bound to the Dynabeads can be stored and reused to screen multiple libraries.

Reader Enquiry No. 115

Atlas

From Clontech

A large-scale mouse cDNA expression array for throughput monitoring of gene expression

The array includes 588 genes involved in key processes, such as oncogenesis, growth and cell-cycle control, apoptosis, embryogenesis and signal transduction. Two arrays are included in each order to facilitate differential expression profiling, for instance, to compare expression levels in transgenic knockout mice to those in normal mice. Because no special equipment is required for analysis, the Atlas mouse array is an economical way to explore the relationships among genes. Arrays are provided on nylon membranes that can be hybridized

ADVERTISEMENTS

DNA SEQUENCING

OLIGONUCLEOTIDES

GENOTYPING

By research labs - For research labs

Call us for more information

Molecular Biology Services
King's College School of Medicine and Dentistry
Tel: +44 171 346 3126
Fax: +44 171 733 3877
www.kcl.ac.uk/links/mbis

READER ENQUIRY NO. 19

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT



PROTEIN SEQUENCING

Membrane-bound, solid or solution samples

£12 per cycle

5 cycles minimum - no set up charge

For further information contact Dr. Joe Gray
Phone: +44 191 222 8612
Fax: +44 191 222 8129
Email: JOE.GRAY@newcastle.ac.uk

READER ENQUIRY NO. 14

Check out our homepage at <http://www.resgen.com>

Peptides
\$18/residue

Antibodies
monoclonal - \$8,500
polyclonal - \$1,195

Complete services for your custom peptide and polyclonal or monoclonal antibody needs.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 29

with a ³²P-labelled first-strand cDNA probe and visualized by autoradiography or phosphorimaging.

Reader Enquiry No. 116

Polymerases and priming

HS TaQuant-Off

From Quantum Biotechnologies

A reagent that blocks thermostable DNA polymerase activity during PCR set-up

This compound can be used with various

native or recombinant enzymes, and its inhibiting activity is completely reversed above 48 °C. According to the manufacturer, pre-mixing solutions of *Taq* DNA polymerase and HS TaQuant-Off do not lead to loss of polymerase activity even after several hours at the room temperature. The company also supplies various thermostable DNA polymerases that are licensed by Hoffman-La Roche for PCR and RT-PCR. Native *Taq* polymerase is a highly purified 94K enzyme isolated from *Thermus aquaticus*. BioTaq is a recombinant enzyme obtained by over expression of the *T. aquaticus* polymerase gene in *Escherichia coli*. Q-BioTaq polymerase is a truncated form of the enzyme, missing the 5'→3' exonuclease activity. It provides enhanced thermostability up to 98 °C and works over a broad range of conditions, facilitating PCR protocol optimization. The low error rate of this enzyme makes it useful for multiplex amplifications, RAPDs and sequencing procedures. The BI-TAQ DNA polymerase is purified from *Thermus brokianus* and is used for generating long amplicons (up to 6 kb) without a complex and time consuming PCR optimization step. This enzyme is suitable for

generating PCR products from GC-rich templates and is said to be preferable to native *Taq* polymerases for sequencing applications as it has a twofold lower error rate.

Reader Enquiry No. 117

Pwo DNA polymerase

From Boehringer Mannheim

A new proof-reading thermostable DNA polymerase

The company claims that this new enzyme product has a lower error rate than any other available on the market, making it well suited to high-fidelity PCR. It amplifies blunt-ended fragments up to a stated 2.9 kb and can therefore be used directly for blunt-end ligation without any pretreatment of the ends, says Boehringer. It exhibits increased thermal stability with a half life of more than two hours at 100 °C. The inherent 3'→5' exonuclease proof-reading activity of *Pwo* results in a stated tenfold increase in fidelity of DNA synthesis compared to *Taq*. The enzyme was originally isolated from the archaeobacterium *Pyrococcus woesei*. Boehringer Mannheim supplies the recombinant form cloned in *Escherichia coli*

and purified to be free of unspecific endo- or exonucleases.

Reader Enquiry No. 118

Primer Premier Mini

From Premier Biosoft

A free program that analyses oligonucleotides to be used in PCR, sequencing or hybridization

With this program users simply type in the sequence of an oligo, and with a single mouse click, the program calculates the melting temperature using nearest neighbour theory, GC percentage, free energy (ΔG) in kcal mol⁻¹, optical activity and degeneracy. In addition, all secondary structures that reduce priming efficiency, such as hairpins and dimers, are shown ranked by structural stability. A primer database makes it easy to store the names and sequences and frequently used primers. The program can be downloaded free from the company's website at <www.PremierBiosoft.com>.

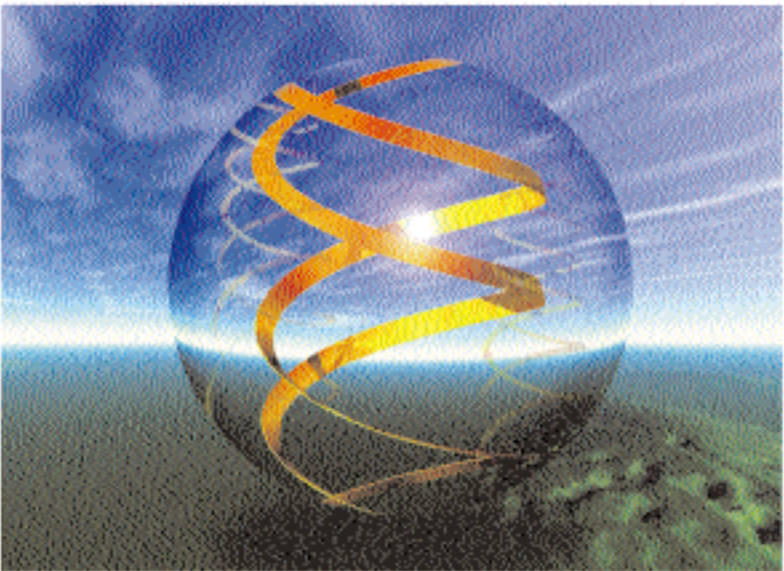
Reader Enquiry No. 119

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

Searching for Something?

Find the right genetic sequence analysis tool for your platform—Windows, Macintosh, UNIX or X-Windows:



DMICA 1.0 for Windows® 95/NT®
Use Project View to organize your sequences, alignments, graphics, presentations and even web site links! Our latest release offers improved publication-quality graphics, added primer design tools and easy access to internal databases.

MacVector™ 6.0 for Macintosh®
Experience a comprehensive, easy-to-use, personal sequence analysis program with outstanding graphics, integrated Internet searches and more.

GCG's Wisconsin Package™ for UNIX and Open VMS
Use the leading enterprise-wide solution for sequence analysis and database searching, with interfaces for accessibility from UNIX, PC or Macintosh systems.

SPS™ products for SWAT, Cross_Match and Phrap
Save hours using these parallel, optimized products for searching and assembly. For example, SPS SWAT™ is 19 times faster on a standard 20-CPU machine.

Call us or visit our web site. Your search is over!
<http://www.oxmol.com/prods/bio>

Desktop Chemistry • Molecular Modeling • Bioinformatics • Databases • Cheminformatics • Collaborations



Oxford Molecular
Solutions for Discovery Research

USA
tel: 800-876-9994
fax: 408-879-6302

Europe
tel: +44 1865 784600
fax: +44 1865 784601

Worldwide
<http://www.oxmol.com/>
1-800-879-6300

©1998 Oxford Molecular. All rights reserved. Oxford Molecular is a registered trademark and the Oxford and GCG Wisconsin Package are trademarks of Oxford Molecular Ltd.
SPS, SWAT, SPS Cross_Match and SPS Phrap are trademarks of Southern Parallel Software. Macintosh is a registered trademark of Apple Computer. Windows and Windows NT are registered trademarks of Microsoft.
Genetics Computer Group (GCG) is located at 535 Science Drive, Madison, Wisconsin 53711 USA. tel: 608-231-5200 fax: 608-231-5202 <http://www.gcg.com/>